Biochemical Characterization of DDX43 (HAGE) Helicase

A Thesis Submitted to the College of

Graduate Studies and Research

In Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Biochemistry

University of Saskatchewan

Saskatoon

By

Tanu Talwar

PERMISSION OF USE STATEMENT

I hereby present this thesis in partial fulfilment of the requirements for a postgraduate degree from the University of Saskatchewan and agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, either in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised this thesis or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts of it for any financial gain will not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Biochemistry
University of Saskatchewan
107 Wiggins Road
Saskatoon, Saskatchewan,
Canada S7N 5E5

ABSTRACT

DDX43, DEAD-box polypeptide 43, also known as *HAGE* (helicase antigen gene), is a member of the DEAD-box family of RNA helicases. It is highly expressed in many tumor types compared with normal tissues, and is therefore considered as a potential target for immunotherapy of cancers. Despite its unique expression and potential as a therapy target, little is known about its biochemical and physiological functions. In this study, we purified recombinant DDX43 protein to near homogeneity and our gel filtration results showed that DDX43 exists as a monomer in solution. Biochemical assays using monomer fractions of DDX43 demonstrated that it could unwind both RNA and DNA substrates in an ATP-dependent manner, and most efficiently in the presence of Mg^{2+}; no significant unwinding activity was detected with other nucleoside triphosphates or divalent cations. Replacing the conserved lysine in motif I (K292A) or aspartic acid in motif II (D396A) abolished the unwinding activity. Intriguingly, DDX43 could unwind RNA substrates without strict polarity, but it showed higher unwinding activity on a 5' tail RNA substrate compared to a 3' tail or blunt end RNA substrates. However, for DNA substrates, it exhibited unidirectional translocation and unwound DNA in a 3'to 5' direction only. A K-homology (KH) domain in the N-terminal region of DDX43 was found to possess strong nucleic acid binding ability and the N-terminal domain showed novel strand exchange and required for the unwinding activity. Compared to the full-length protein, the C-terminal helicase domain had weaker unwinding activity, but an increase was observed in the presence of the N-terminal domain. Using Co-immunoprecipitation and mass spectrometry, we found that DDX43 associates with pICln and MEP50, two methylosome subunits that are involved in assembly of the spliceosome complex. Collectively, our results suggest that DDX43 is a KH domain containing ATP-dependent dual helicase, where unwinding activity is mediated through cooperation between its N-terminal domain and helicase domain, and potentially involved in pre-mRNA splicing.

ACKNOWLEDGEMENTS

I wish to express the deepest appreciation to my supervisor Dr. Yuliang Wu for his endless support on academic studies and his helpful guidance on my life. His resolute enthusiasm for science kept me constantly engaged with my research. I would also like to express gratitude to my committee member Dr. Jeremy Lee, Dr. Stanley Moore and Dr. Erique K. Lukong for their insightful suggestions and guidance throughout my project. My sincere thanks also go to everyone in Dr. Wu's lab, including Dr. Manhong Guo, Dr. Venkatasubramanian Vidhyasagar, Hao ding and Ahmad Kareim for their stimulating discussions. I would like to thank all members of the Cancer Cluster and faculty of Biochemistry Department. I must express my profound gratitude to my parents, my brother and my In-laws for providing me with continuous encouragement and for their patience throughout my years of research. And finally, I acknowledge the love of my life, my husband, Vishal Taimni, who blessed me with a life of joy and always encouraged me to pursue my dreams. I am also thankful to all my friends for their help and support through my master program.

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABCB5+	ATP-binding cassette subfamily B member 5
ADP	Adenosine diphosphate
AML	Acute myeloid leukemia
AMP-PNP	Adenylyl-imidodiphosphate
ATP	Adenosine triphosphate
ATP γ S	Adenosine 5′ -[γ -thio] triphosphate
bp	Base pair
BSA	Bovine serum albumin
CML	Chronic myeloid leukemia
CTA	Cancer testis gene
CV	Column volume
DMEM	Dulbecco's modified eagle's medium
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assays
ER	Estrogen-receptor
FBS	Fetal bovine serum
FMRP	Fragile X mental retardation protein
FPLC	Fast protein liquid chromatography
HAGE	Helicase antigen gene
HD	Helicase domain
His	Hexa-histidine
hnRNP	K heterogeneous nuclear ribonucleoprotein K
HRP	Horseradish peroxidase
IHC	Immunohistochemistry

IPTG	Isopropyl β-D-1-thiogalactopyranoside
JAK-STAT	Janus kinase–signal transducers and activators of transcription
KH	K-homology
LB	Lysogeny broth
MDS	Myelodysplastic syndrome
MEK	Mitogen-activated protein kinase
MEP50	Methylosome protein 50
MOPS	3-(N-morpholino) propanesulfonic acid
NE	No enzyme
Ni-NTA	Nickel-nitrilotriacetic acid
NRAS	Neuroblastoma RAS
NTP	Nucleotide triphosphate
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEI	Polyethylenimine
Pi	Inorganic phosphate
PMSF	Phenylmethanesulfonylfluoride
PML	Progressive multifocal leukoencephalopathy
PVDF	Polyvinylidene difluoride
Sam68	Src-Associated substrate in mitosis of 68 kDa
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SF	Superfamily
TBE	Tris/Borate/EDTA
TCEP	Tris (2-carboxyethyl) phosphine

TEMED	N,N,N’,N’ – tetramethylethylenediamine
TLC	Thin-layer chromatography
Tris	Tris (hydroxymethyl) aminomethane
TNBC	Triple-negative breast cancer
WT	Wild type

1. INTRODUCTION

1.1 Helicases and their Functions

Helicases are ubiquitous and integral members of almost all complexes that catalyze reactions of nucleic acid metabolism. The basic reaction catalyzed by this family of enzymes is the unwinding of the duplex form of DNA or RNA, a process coupled to the steps of enzymatic nucleoside triphosphate (NTP) hydrolysis (Caruthers and McKay, 2002). Based on their substrates, helicases can be classified as DNA or RNA helicases, although some can function on both DNA and RNA molecules (Pyle, 2008; Taylor *et al*., 2010). Generally helicases move directionally along one nucleic acid strand and can be classified according to their directionality of unwinding, 5'–3' or 3'–5' with respect to the strand they bind to and move along (Patel and Picha, 2000), but they can catalyze bidirectional unwinding as well (Tuteja *et al*., 2014).

Both DNA and RNA helicases are involved in regulation of various aspects of nucleic acid metabolism (Bernstein *et al.*, 2010; Jankowsky, 2011). DNA helicases have been reported to function in a variety of biochemical processes, including unwinding duplex and alternative DNA structures (triplex, G-quadruplex) and displacing protein bound to DNA (Wu and Brosh, Jr., 2012; Suhasini and Brosh, Jr., 2013). They are also involved in other metabolic processes, e.g., DNA replication, meiotic recombination, DNA double-strand break repair and telomere maintenance (Singh *et al*., 2012). DNA helicases also play multifaceted roles in homologous recombination (HR) and chromatin remodeling (Daley *et al*., 2013).

The functions of RNA helicases are not limited to RNA strand separation, as they can also catalyze various RNA chaperone activities. They can displace proteins from RNA molecules without duplex unwinding, anneal RNA strands, act as RNA clamps or placeholders, and stabilize on-pathway folding intermediates and thus regulate RNA metabolism (Fairman *et al*., 2004; Halls *et al*., 2007; Shibuya *et al*., 2004; Solem *et al.*, 2006; Yang and Jankowsky, 2005). DEAD-box proteins are members of the SF2 superfamily of helicases, comprised of mainly RNA helicases and are involved in virtually all aspects of

eukaryotic RNA metabolism. Prp5 (DDX46) (Xu *et al*., 2004) and p68 (DDX5) (Liu and Cheng, 2015) are implicated in the early events of pre-splicesomal assembly. A large number of DEAD-box proteins are associated with rRNA maturation in eukaryotes and prokaryotes during ribosome biogenesis and comprises by far the largest number of RNA helicases of the DEAD box family (Liang and Fournier, 2006). For example, two DEAD-box proteins, eIF4A and Ded1 from *S. cerevisiae* and their homologues in higher eukaryotes, are essential for translation initiation (Linder, 2003). Some DEAD-box proteins are also associated with transcription (p68 and p72) (Fuller-Pace, 2006), nuclear export (Dbp5) (Gatfield *et al*., 2001), RNA export and RNA degradation (Dob1/Mtr4) (Anderson and Parker, 1998; Estruch and Cole, 2003), and in organelle gene expression (Mss116, Mrh4, Cyt-19) (Schmidt *et al*., 2002).

1.2 The Mechanisms of Helicases

Helicases unwind long stretches of duplex nucleic acids by coupling base pair separation to translocation to provide the single-stranded template required in many biological events, such as replication, recombination, and repair (Soultanas and Wigley, 2001). To unwind substrates most of the helicases need a single-stranded nucleic acid region to bind and to initiate their action of strand separation. Once loaded on the strand, they show a directional bias and translocate either 5’ to 3’ or 3’ to 5’ direction or may unwind DNA in a bidirectional manner (Tuteja *et al*., 2014). But a few helicases like RecBCD and DEAD-box helicases can initiate unwinding from blunt ended substrates as well (Farah and Smith, 1997). All of these mechanisms involve NTPase-coupled nucleic acid affinity changes and a conformational change to explain biased movement that results in base pair separation or translocation (Patel and Donmez, 2006). To explain the key features of the unwinding process, two catalytic mechanisms have been proposed referred to as active and passive. (Manosas *et al*., 2010).

In the passive mechanism, the helicase behaves opportunistically. The Helicase does not interact with the double strand DNA, it interacts only with single-stranded DNA (ssDNA) when terminal base pairs open, which occurs rapidly and spontaneously at duplex termini.

The Protein binds to the ssDNA, stabilizes it and physically blocks the base pair(s), and prevents immediate reannealing of the duplex **(Fig. 1.1 A)**. Because the terminal base pair at the junction opens and closes at a very fast rate, this type of a mechanism is attractive for helicases that can move and occupy one base at a time. Also, this mechanism of unwinding is very efficient because the kinetics of local base pair opening are relatively fast (Nonin *et al*., 1995). For example, Mcm 4,6,7 (30 bp) , NPH II (40 bp), unwind substrates with high processivity using this mechanism (Kaplan *et al*., 2003; Kawaoka *et al*., 2004; McGeoch *et al*., 2005; Hacker and Johnson, 1997)

When the rate constants for base-pair opening are slower than the rate constants for helicase translocation along single-strand RNA or DNA, helicases use the active mechanism for the unwinding of the substrate. In this mechanism, the helicase binds to the duplex and induces strand separation, an event directly coupled to its movement along the DNA and fully separates the strands in a unidirectional manner. The protein could spread out over the duplex region and electrostatically destabilize it **(Fig. 1.1 B)**. Enzymes using this type of mechanism would translocate on ssDNA and unwind the DNA duplex at a rate comparable or identical to that of translocation on ssDNA (Bianco and Webb, 2012). NS3 (Cheng *et al*., 2007), hepatitis C virus (non-ring shaped helicase) (Donmez *et al.*, 2007), and Pcr A helicase (Cheng *et al*., 2002) have been proposed to use this mechanism for unwinding.

1.3 Superfamily 2 Helicases

Based on the occurrence and characteristics of conserved motifs in the primary sequence, helicases are categorized into six superfamilies, designated SF1–SF6 **(Fig. 1.2)**. Helicases of different families share similarities in their three-dimensional folds, with two RecA-like domains making up the helicase core (Subramanya *et al*., 1996; Bird *et al*., 1998). Structurally, two distinct types of helicases exist, those forming toroidal, predominantly hexameric structures, and those that do not. The toroidal enzymes comprise SF3 to SF6, and the non-ring forming ones comprise SF1 and SF2 (Singleton *et al*., 2007). The largest of the

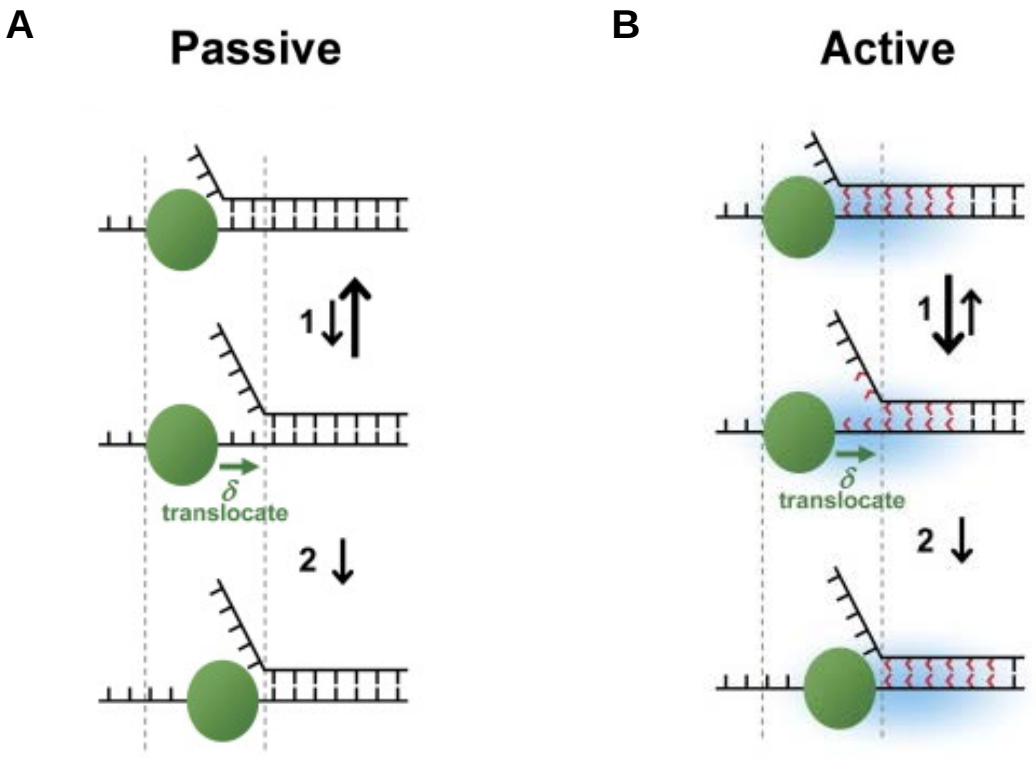


Figure 1.1 The mechanisms for unwinding by helicases. (A) In the passive mechanism the DNA fork thermally fluctuates between dsDNA and ssDNA states (step 1). When the amount of ssDNA between the helicase and the junction is greater than or equal to the helicase step size (δ) the helicase may forward translocate (step 2). **(B)** In active mechanism helicase destabilizes a region (the light blue cloud) of dsDNA near the junction (step 1). This makes the junction more likely to be open so that the helicase is able to step forward (step 2) more frequently. Adapted from Pyle AM, 2008.

six groups are the superfamily 1 and 2 helicases, most of which have a 3'-5' directionality. In humans, there are at least 103 SF2 and 17 SF1 helicases, in *S. cerevisiae* at least 59 SF2 and 9 SF1 helicases have been identified. All SF1 and SF2 helicases contain Walker A and Walker B motifs, and typically five to seven different flanking conserved motifs. Walker A motif, also known as the Walker loop or P-loop (phosphate-binding loop) has the consensus sequence GxxxxGK(S/T), which is associated with phosphate binding (Mercier *et al*., 2015). Helicase motifs involved in ATP binding are located at the interface between two RecA-like domains (Singleton and Wigley, 2002). The two helicase domains are connected through a linker, which varies considerably between SF1 and SF2 proteins and also between SF2 family

members (Caruthers and McKay, 2002). With bound ATP and nucleic acid, the two helicase domains form a deep cleft. The ATP binding site is located on one side of this cleft, and the nucleic acid binds on the other side, across both domains.

All superfamilies of helicases from SF1 to SF6 show variation in terms of conserved motifs and their members. In terms of number of conserved motifs, SF1 and SF2 contain two RecA-like domains and at least seven conserved amino acid motifs whose sequences, organization, and secondary structures are in general very similar (Gorbalenya *et al*., 1988). SF1 (Rep and PcrA) and SF2 (NS3) helicases differ primarily in motifs III and IV (Kim *et al*., 1998). Members of SF3 to SF6 superfamilies are structurally different from SF1 and SF2 helicases as they contain only one RecA-like domain and only two to five conserved motifs, e.g., SF3 and SF4 contains only 4 and 5 conserved sequence motifs respectively. In terms of the subfamilies and members of the different superfamilies, SF2 includes several intensively studied subfamilies, including DEAD-box RNA helicases (Cordin *et al*., 2006), the RecQ-like family (Bennett and Keck, 2004), and the Snf2-like enzymes (Flaus and Owen-Hughes, 2004). SF3 enzymes, usually derive from DNA or RNA viruses, e.g., SV40 and AAV. SF4, and SF5 include DNA helicases of bacterial origin, e.g., the transcription terminator Rho/V-F-ATPases (Enemark and Joshua-Tor, 2008). MCM proteins of archaea and eukaryotes and RuvB branch migration enzyme belong to the SF6 (Putnam *et al.*, 2001).

The superfamily 2 is subdivided into at least 10 families, based on phylogenetic analysis of the sequences of the helicase core domains (Jankowsky and Fairman, 2007). Five of these SF2 families (DEAD-box, DEAH/RHA, Ski2-like, RIG-I-like, and viral DExH proteins includes NS3/NPH-II family) are comprised mainly of RNA helicases and are thus termed ‘‘RNA helicase families’’. Out of five, three of them, DEAD box, DEAH box, and Ski2 (Superkiller-like 2)-like share eight conserved motifs (I, Ia, Ib, II, III, IV, V and VI) and are collectively referred to as the DExD/H group of RNA helicases (Jankowsky and Bowers, 2006). These conserved motifs occupy two different domains: motifs I, Ia, Ib, II, and III in domain 1 (Rec A1) and motifs IV, V, and VI in domain 2 (Rec A2). These motifs are usually

clustered in a region of 200–700 amino acids, called the helicase core domain.

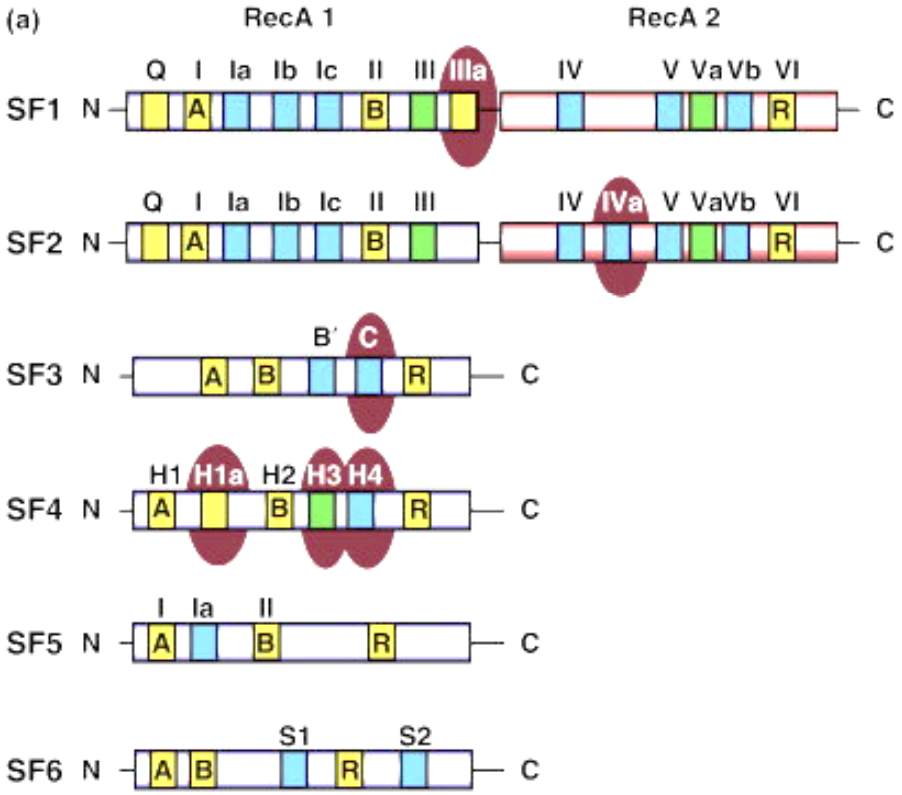


Figure 1.2 A schematic representation of the core helicase domains of helicase superfamilies. The N-terminal RecA domain (RecA1) is represented by a blue cylinder and the C-terminal RecA domain (RecA2) is shown as a red cylinder. Conserved amino acid motifs are colored according to helicase function. Motifs in yellow are involved in NTP binding/hydrolysis, green are associated with translocation, and blue interact with the nucleic acid. Motifs that are unique to specific superfamilies are highlighted with a red oval. The Walker A (A), Walker B (B) and arginine finger (R) motifs are conserved across all helicase superfamilies. Adapted from Jackson RN *et al.*, 2014.

These conserved motifs are important for biochemical functions of helicases. Both motifs I and II are required for binding and hydrolyzing ATP. Mutations in either motif I or II result in enzymes that are deficient in ATPase and DNA/RNA-unwinding activities (Cordin *et al.*, 2006). Motif II contains a conserved aspartate residue that is thought to bind the Mg^{2+} ion, which is required for catalysis and it also contains a conserved glutamine residue that is

proposed to coordinate the attacking water during ATP hydrolysis (Story *et al*., 2001). Motif III couples ATP hydrolysis with the helicases activity and motif VI has a role in RNA binding and ATP hydrolysis (Pause *et al.*, 1993). Motifs Ia, Ib, IV, and V are less well studied, but there is evidence that they contribute to substrate binding (Cordin *et al.*, 2006) **(Fig. 1.3 A)**. X-ray crystallographic studies demonstrate that the conserved helicase motifs are closely associated with the tertiary structure of helicase proteins, suggesting that they form a large functional domain coordinating ATP binding and hydrolysis to nucleic acids unwinding (Singleton *et al*., 2007) **(Fig. 1.3 B)**.

1.4 DEAD-box Helicases

The DEAD-box proteins were first identified as a distinct family in the late 1980's when alignments based on eight homologues of the yeast eIF4A translation initiation factor highlighted the presence of several conserved motifs (Linder *et al.*, 1989). The name of the family was derived from the amino-acid sequence D-E-A-D (Asp–Glu–Ala–Asp) of its Walker B motif. This key motif of DEAD-box proteins have been found in more than 500 proteins (Silverman *et al*., 2003).

DEAD-box proteins are involved in virtually all aspects of eukaryotic RNA metabolism. Several DEAD-box helicases have generated particularly high interest because they are involved in the multiple processes of RNA metabolism (Fuller-Pace, 2013). Dysregulation of RNA helicases can influence various critical processes, namely transcription, splicing, and translation, which may all lead to cancer (Abdelhaleem, 2004).

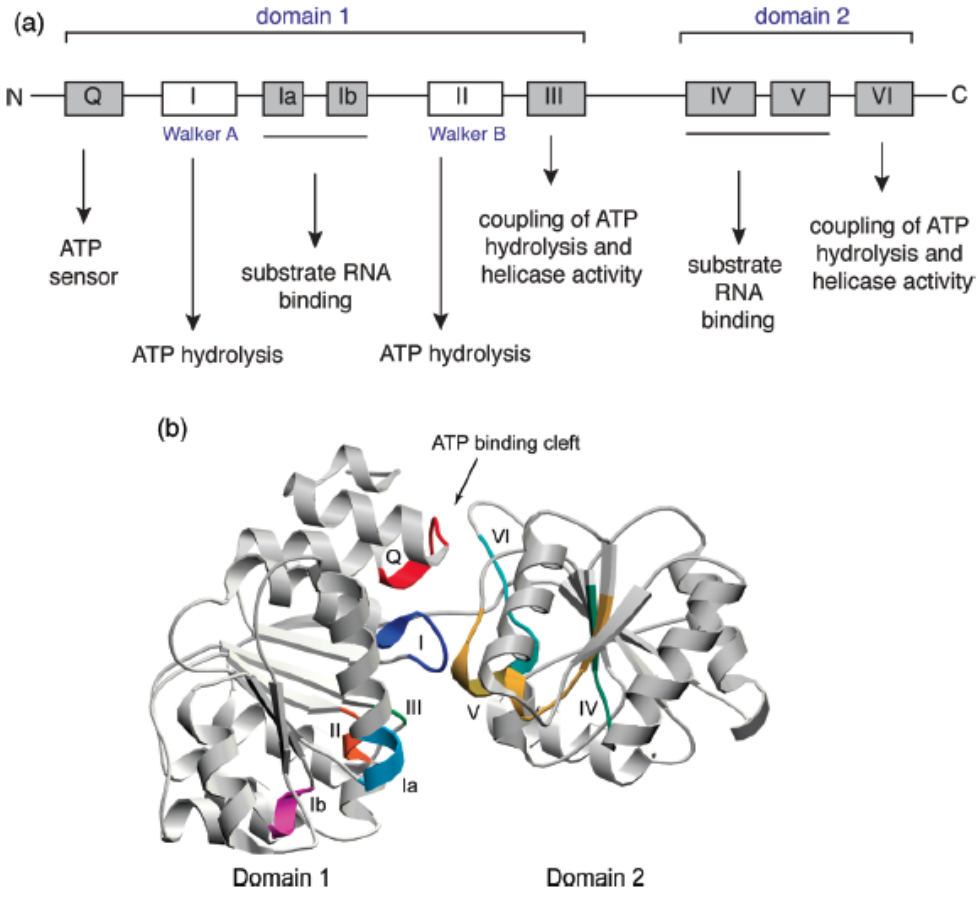


Figure 1.3. Domain architecture and X-ray structure of the SF2 superfamily helicase. **(A)** Schematic representation of the motifs found in the SF2 superfamily. The motif II (Walker B) contains the DEAD (named after the amino acids Asp-Glu-Ala-Asp) box. Adapted from Bleichert *et al*., 2007. **(B)** Crystal structure of the *Methanococcus jannashii* DEAD box protein. The conserved motifs shown in are highlighted with different colors. The figure was created with molscript (http://www.avatar.se/molscript/) and POVRAY (http://www.povray.org/) using the coordinates deposited in the Protein Data Bank (1HV8).

RNA helicases which are generally overexpressed in various cancers include DDX5 (colorectal cancer), DDX47 (cervical cancer), DDX48 (gastric cancer) and DHX9 (lung cancer) (Causevic *et al*., 2001; Doorbar *et al*., 2000; Lee *et al*., 2002; Wei *et al*., 2004). DEAD-box helicases have also been implicated in tumor initiation, progression and maintenance (Robert and Pelletier, 2013). These proteins generally act as components of multi-protein complexes and regulate cellular proliferation and/or neoplastic transformation

(Fuller-Pace, 2013).

DEAD-box proteins have been found to unwind their RNA substrates by a unique mechanism. Duplex unwinding by DEAD-box proteins is not based on translocation of the RNA substrate (Yang and Jankowsky, 2006). Instead, DEAD-box proteins directly load to the duplex region, and then pry the strands apart by local destabilization in an ATP-dependent fashion (Yang *et al*., 2007). As a result, DEAD-box proteins do not display a defined unwinding polarity on an RNA substrate, as distinct from many other RNA and DNA helicases.

1.5 The KH domain

The KH domain is an evolutionarily conserved sequence motif that was first identified in the human heterogeneous nuclear ribonucleoprotein K (hnRNP K) more than two decades ago (Matunis *et al.*, 1992). KH domains are approximately ~70 amino acids long with the most conserved consensus sequence being VIGXXGXXI, mapping to the middle of the domain (Gibson *et al*., 1993) **(Fig. 1.4 A).** The K homology (KH) module is a widespread RNA-binding motif that has been detected by sequence similarity searches in proteins as hnRNP K and ribosomal protein S3 (**Fig. 1.4 A**). Analysis of three-dimensional structures of KH domains in hnRNP K (Baber *et al*., 1999) and S3 (Wimberly *et al*., 2000) reveals that they are topologically dissimilar and thus belong to different protein folds . Thus KH motif proteins provide a rare example of protein domains that share significant sequence similarity in the motif regions but possess globally distinct structures (Grishin, 2001). Sequence similarity in the KH motif reflects descent from a common ancestor (Burd and Dreyfuss, 1994). The conservation of KH motif in diverse organisms such as Bacteria, Archaea and Eukaryotes suggests that KH arose early in evolution.

The KH domain exists in two different versions; type 1 KH fold (KH1) in eukaryotes and type 2 KH fold (KH2) in prokaryotes (Ashley, Jr. *et al.*, 1993). In eukaryotes, it is present in multiple copies, and can vary from 2 to 15 copies of the KH domain (Gibson *et al*., 1993;

Siomi *et al.*, 1994). In prokaryotes, the KH domain is present in single copy, for example in Mer1p and Sam68 (Spingola *et al*., 2004; Lukong and Richard, 2003). The presence of multiple copies of the KH domain is a sign that it binds RNA as a multimer. Type I and type II KH domains share a minimal consensus sequence motif and contain different structural folds with different C-terminal or N-terminal extensions giving distinct topology for type I and type II, respectively (Valverde *et al*., 2007) . The type I KH domains have a three stranded anti-parallel beta-sheet packed against 3 alpha helices whereas in type II domains two of the three beta strands are in a parallel orientation **(Fig. 1.4 B)**.

Unlike RNA recognition motifs, which recognize a diversity of RNA lengths, the binding cleft of KH domains is versatile but accommodates only four nucleic acid bases. In all KH domain-nucleic acid structures, the nucleic acid backbone interacts with a conserved GxxG loop that links the two helices of the minimal KH core. The RNA or DNA is bound in an extended, single stranded conformation across one face of the KH domain, between the α1-helix and the α2-helix and GXXG on the 'left', and the β2-sheet and the variable loop on the right, mainly by hydrogen bonds **(Fig. 1.4 C)**. This orients four nucleic acid bases towards a groove in the protein structure where hydrophobic interactions and a network of main-chain and side-chain hydrogen bonds mediate nucleobase recognition (Nicastro *et al*., 2015a). The phosphates of the first two nucleotides are instead docked against the GxxG loop by means of electrostatic interactions, H-bonding and inter-molecular Van der Waals interactions, depending on the specific complex. Nucleobase recognition is mediated by base-pair-like H-bonding between the moieties on the Watson–Crick edges of the NA bases and the backbone and side chain of the protein (Backe, Messias *et al*. 2005)

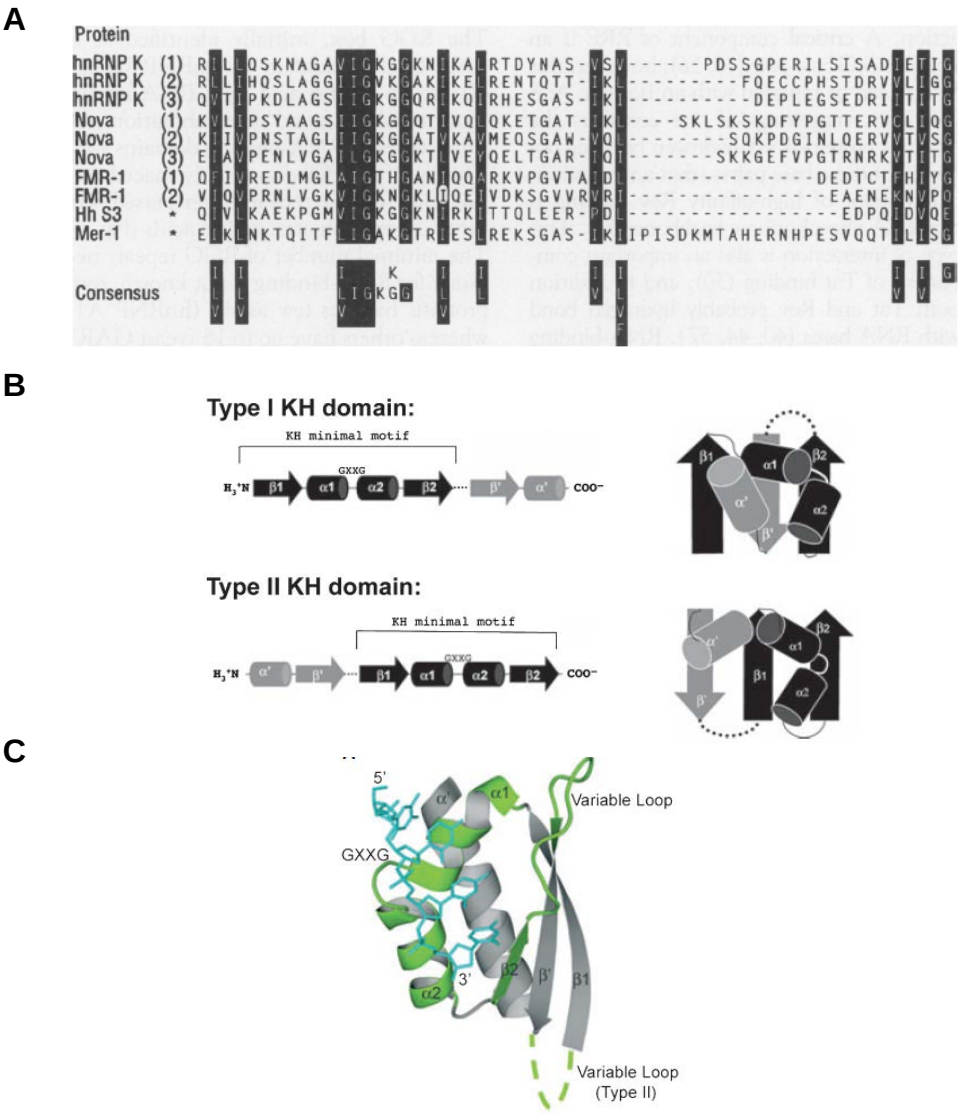


Figure 1.4 hnRNP K homology (KH) motif and crystal structure of a KH domain bound to RNA. **(A)** Sequence alignment of the hnRNP K motif across species. **(B)** Domain representation of type I and type II KH domain folds. The dotted line connecting the β2-strand and β'-strand represents the variable loop. The white line connecting the α1-helix and the α2-helix represents the GXXG loop. **(C)** 3D representation of type I KH domain; the binding cleft comprises the secondary structural elements α1-helix, GXXG loop, α2-helix, β2-strand, and variable loop (colored green), and recognizes four nucleotides (cyan sticks). The green dotted line represents the location of the variable loop in type II KH domains. In all KH domain-nucleic acid structures, the nucleic acid backbone interacts with a conserved GxxG loop that links the two helices of the minimal KH core. The nucleic acid is bound in an extended conformation across one side of the KH domain mainly by hydrogen bonds. This orients four nucleic acid bases towards a groove in the protein structure where hydrophobic interactions and a network of main-chain and side-chain hydrogen bonds mediate nucleobase recognition. Adapted from Burd and Derfuss, 1994 and Valverde *et al*., 2008.

In eukaryotes, RNA-binding proteins that contain multiple KH domains play a key role in coordinating the different steps of RNA synthesis, metabolism, and localization (Siomi *et al.,* 1994*).* Sequence and structure-specific recognition of RNA targets is a hallmark of proteins involved in pre-messenger RNA splicing and other kinds of RNA processing (Burd and Dreyfuss, 1994). KH domains in Nova2 have been found to regulate alternative splicing of pre-mRNA, having a specific function in the brain (Ule *et al*., 2005). Thus hnRNP K-homology (KH) domain has emerged as one of the most prevalent protein motifs responsible for ssRNA binding in these systems.

1.6 DDX43 (HAGE) Helicase

DDX43, DEAD-box polypeptide 43, also known as HAGE (helicase antigen gene), was first identified as a cancer/testis antigen gene in a human sarcoma cell line (Martelange *et al*., 2000). DDX43 belongs to the DEAD-box family of ATP-dependent RNA helicases and the gene is present on chromosome 6 (6q12-q13), which was determined by radiation hybrid analysis, and encodes a putative 73 kDa protein. Close inspection of the DDX43 protein reveals that it contains the signature motifs of a SF2 RNA helicase, including Q, I, Ia, Ib, II, III, IV, V, and VI **(Fig. 1.5A).** In addition, DDX43 also has conserved motifs Ic, Va, and Vb and a KH domain with its hallmark sequence of VIGXXGXXI which is present in certain DEAD-box proteins (Fairman-Williams *et al*., 2010). **(Fig. 1.5 B)**.

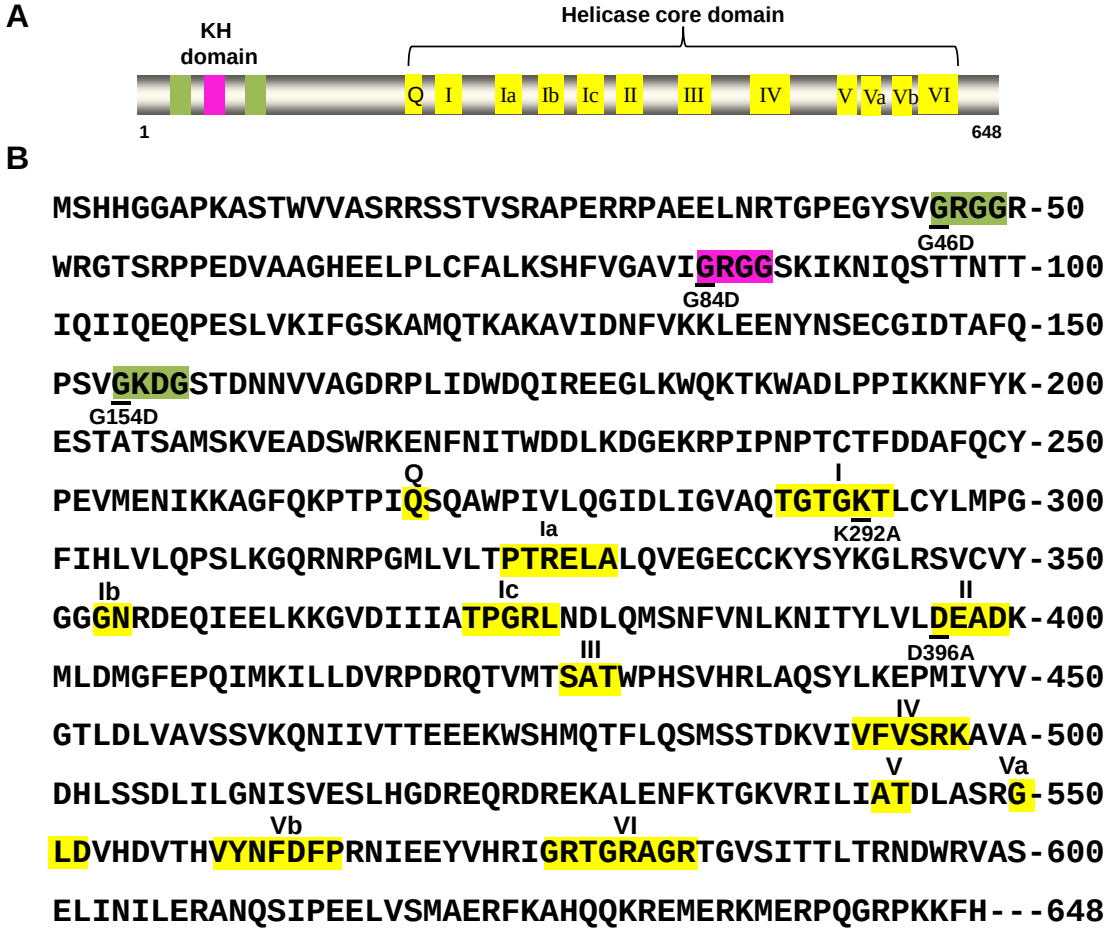


Figure 1.5 Domains and primary protein sequence of the human DDX43 protein. (A) Motifs of SF2 helicase DDX43, Conserved helicase motifs are indicated by yellow boxes, and three potential K homology (KH)-domain are indicated with pink. **(B)** Schematic representation of the primary structure of DDX43.

DDX43 mRNA is found to be expressed by a wide range of tumour tissues at levels at least 100-fold that of normal tissues, with the exception of the testis (Martelange *et al*., 2000) **(Fig. 1.6 A).** At the protein level, DDX43 is also detected in a variety of tumour tissues including bladder, brain, breast, colon, esophagus, kidney, liver, lung, stomach and small intestine, but totally absent or expressed at very low level in normal tissues (Mathieu *et al*., 2010) **(Fig. 1.6 B).** DDX43 protein is also overexpressed in more than 50% of chronic myeloid Leukemia (CMLs), 20% of acute myeloid Leukemia (AMLs) (Adams *et al*., 2002), and more than 40% of multiple myelomas. It is also highly expressed in melanoma (Mathieu *et al.*, 2007). DDX43 has been detected in 1650 breast cancer patients, where high expression

correlates significantly with aggressive clinicopathological parameters, including high proliferation, overexpression of both EGFR (epidermal growth factor receptor) and HER2 (human epidermal growth factor receptor 2), and abnormal expression of p53 and p16 (Wiese and Pajeva, 2014). Therefore, DDX43 is proposed to be a valid candidate for designing a broad spectrum vaccine against cancer. Considering the expression pattern and the diversity of DDX43 expression in different tumors, DDX43 may also represent a suitable target for immunotherapy.

DDX43 belongs to superfamily 2 RNA helicases, with seven conserved helicase motifs. RNA helicases are involved in gene transcription and post-transcriptional regulation, including pre-mRNA splicing, translation initiation/elongation, and RNA degradation (Cordin *et al.*, 2006). Their altered expression levels have also been implicated in tumour initiation, progression, and maintenance (Robert and Pelletier, 2013). DDX43 functions as an oncogene in various biological processes. It has been found to promote the expression of Neuroblastoma RAS (NRAS), an oncogenic protein, by unwinding mRNA secondary structures and enhancing NRAS translation (Linley *et al.*, 2012). It also increases the expression and activation of RAS in uveal melanoma cells with the GNAQ mutation, and mediates resistance to MEK (mitogen-activated protein kinase) inhibitors (Ambrosini *et al.*, 2014) DDX43 has been found to be involved in the transcriptional repression of the tumour suppressor gene PML (promyelocytic leukemia), by inhibiting the activation of JAK–STAT (janus kinase–signal transducers and activators of transcription) pathway through the suppression of cytokine signaling 1 (SOCS1) in ABCB5+ (ATP-binding cassette subfamily B member 5) malignant melanoma-initiating cells (Mathieu *et al.*, 2014)., thereby preventing anti-proliferative and pro-apoptotic functions of PML.

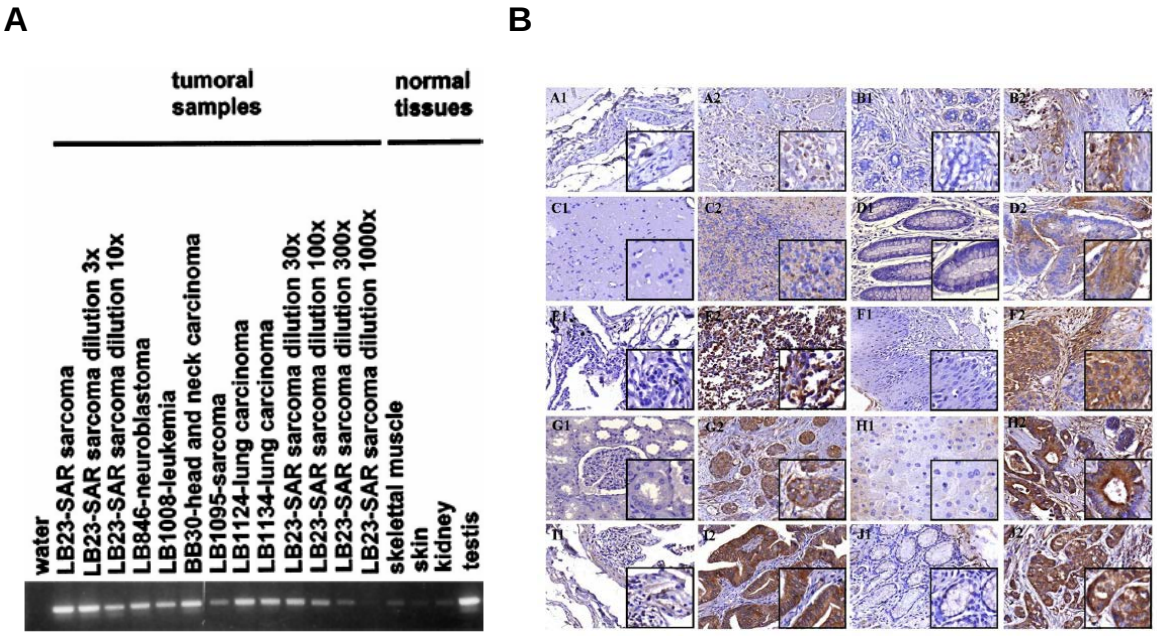


Figure 1.6 High-level expression of DDX43 in tumors. (A) RT-PCR analysis of DDX43 gene expression in normal tissue and tumors. **(B)** DDX43 protein expression in multiple cancer tissue microarrays and patient-matched normal tissues determined by immunohistochemistry. Immunohistochemical staining demonstrated the *in vivo* expression of DDX43 protein at a low level in bladder transitional cell carcinoma (A2) and breast invasive ductal carcinoma (B2); at an intermediate level in astrocytoma (C2), colon adenocarcinoma (D2) and lung squamous cell carcinoma (E2); at a high-level in esophagus small cell carcinoma (F2), kidney clear cell carcinoma (G2), hepatocellular carcinoma (H2), small intestine papillary adenocarcinoma (I2) and stomach adenocarcinoma (J2) but not in their respective matched normal tissues (A1, B1, C1, D1, E1, F1, G1, H1, I1 and J1). Adapted from Martelange *et al.*, 2001 and Mathieu *et al.*, 2010.

Thus, HAGE prevents the activation of JAK–STAT pathway responsible for the induction of the expression of the tumour suppressor PML. Altogether, these observations indicate the role of DDX43 as an oncogene. DNA hypomethylation plays a significant role in tumorigenesis. A wide range of cancer-linked hypomethylation of the genome has been found in almost all of the major human tumors, such as colon, gastric, lung, liver, breast, bladder, and ovarian cancers (Pogribny and Beland, 2009). In a recent study, *DDX43* has been found to be differentially methylated in schizophrenia and considered as an epigenetic marker of the

disease (Wockner *et al.*, 2015). *DDX43* promoters are also found to be hypomethylated in AML patients and thus DDX43 may be related to the pathogenesis of AML. In AML and MDS (Myelodysplastic syndrome), *DDX43* hypomethylation was associated with a favorable outcome and better survival (Lin *et al.*, 2014; Chen *et al*., 2012). In CML patients, abnormal hypomethylation of the *DDX43* promoter was present, which was associated with DDX43 mRNA overexpression and disease progression (Roman-Gomez *et al*., 2007). During the development and progression of acute leukemia, *DDX43* may be activated by promoter hypomethylation and inhibit the proliferation of specific leukemic cells.

Altogether, the emerging picture of DEAD-box helicases depicts a large family of proteins which play an important role in RNA and DNA metabolism. Also, considering its almost unique and specific expression in tumours, DDX43 represent one of the most promising groups of tumour-associated antigens identified to date. A remarkable range of ATP-dependent and ATP-independent activities has been reported for DEAD-box proteins, but no KH domain has been reported in any helicase so far. Thus, the knowledge of their biochemical activities is essential for devising physical models depicting DDX43's functions.

2. HYPOTHESIS AND OBJECTIVES

2.1 Hypothesis: We hypothesize that DDX43 is an ATP-dependent RNA helicase and contains a KH domain in the N-terminal that might function as assessory domain and play a crucial role in biochemical activities of the protein.

2.2 Objectives:

1) To characterize the enzymatic activities of recombinant DDX43 protein, including ATPase, nucleic acid binding, unwinding, and annealing activity.
2) To determine the role of the KH domain in DDX43, and its functional interaction with the helicase domain.
3) To examine the expression and localization of DDX43 in breast cell lines and breast tissues.
4) To determine DDX43's associated proteins by immunoprecipitation and mass spectrometry.

3. MATERIALS AND METHODS

3.1 Reagents

Table 3.1 List of reagents, catalog number and suppliers

Reagents	Suppliers	Address
2-log DNA ladder, N3200L	Biolab	Lawrenceville, Georgia, USA
3XFLAG peptide	APExBIO	Houston, Texas, USA
Acrylamide, 0314	AMRESCO	North York, Ontario, Canada
N,N'-Methylenebisacrylamide	Fisher Scientific	Madison, Wisconsin, USA
Adenosine Triphosphate (ATP), A1852	Sigma-Aldrich	Oakville, Ontario, Canada
Adenosine Diphosphate (ADP), A2754	Sigma-Aldrich	Oakville, Ontario, Canada
Adenylyl-imidodiphosphate (AMP-PNP), A2647	Sigma-Aldrich	Oakville, Ontario, Canada
Adenosine 5′-[γ-thio] triphosphate (ATP-γ-S)	Enzolifesciences	Brockville, Ontario, Canada
Agarose， 5510UB	Gibco	Burlington, Ontario, Canada
Ampicillin, A1593	Sigma-Aldrich	Oakville, Ontario, Canada
Ammonium persulfate (APS), A3678	Sigma-Aldrich	Oakville, Ontario, Canada
Boric Acid, SLBC5554V	Sigma-Aldrich	Oakville, Ontario, Canada
Bovine serum albumin	Sigma-Aldrich	Oakville, Ontario, Canada
Bradford protein assay reagent, 500-0013	Bio-Rad	Hercules, California, USA
Bromophenol blue	Sigma-Aldrich	Oakville, Ontario, Canada
Chloramphenicol	Fisher Scientific	Madison, Wisconsin, USA
Dithiothreitol (DDT)	Sigma-Aldrich	Oakville, Ontario, Canada
DNAase	Sigma-Aldrich	Oakville, Ontario, Canada
dNTPs, N0447S	NEB	Mississauga, Ontario, Canada
Dulbecco's Modified Eagle Medium (DMEM), SH30022.01	Sigma-Aldrich	Oakville, Ontario, Canada
EDTA	Fisher Scientific	Madison, Wisconsin, USA
ANTI-FLAG® M2 Affinity Gel	Sigma-Aldrich	Oakville, Ontario, Canada
Fetal Bovine Serum (FBS), F6178	Sigma-Aldrich	Oakville, Ontario, Canada
Formic Acid	Sigma-Aldrich	Oakville, Ontario, Canada
Glycine, BP381-5	Fisher Scientific	Madison, Wisconsin, USA
Glycerol, 123170	Fisher Scientific	Madison, Wisconsin, USA

Reagents	Suppliers	Address
Imidazole	Fisher Scientific	Madison, Wisconsin, USA
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Fisher Scientific	Madison, Wisconsin, USA
Kanamycin, 60615	Sigma-Aldrich	Oakville, Ontario, Canada
Lithium chloride (LiCl)	Sigma-Aldrich	Oakville, Ontario, Canada
Lysogeny Broth (LB) with agar, L2897	Sigma-Aldrich	Oakville, Ontario, Canada
Methanol, 154246	Fisher Scientific	Madison, Wisconsin, USA
Magnesium Chloride (MgCl2)	Sigma-Aldrich	Oakville, Ontario, Canada
3-(N-morpholino) propanesulfonic acid (MOPS)	Sigma-Aldrich	Oakville, Ontario, Canada
Nickel-NTA Affinity beads	Sigma-Aldrich	Oakville, Ontario, Canada
Non-fat dry milk, 170-6404	Bio-Rad	Hercules, California, USA
Nonidet P-40, 74385	Sigma-Aldrich	Oakville, Ontario, Canada
N,N,N',N'-Tetramethylethylenediamine (TEMED), 87689	Sigma-Aldrich	Oakville, Ontario, Canada
Penicillin-Streptomycin, 084M4778V	Sigma-Aldrich	Oakville, Ontario, Canada
Phenylmethylsulfonyl fluoride (PMSF), P7626	Sigma-Aldrich	Oakville, Ontario, Canada
Protein standard, 161-0374	Bio-Rad	Hercules, California, USA
Protease inhibitor, 05892791001	Roche	Mannheim, Germany
Polyvinylidene difluoride (PVDF)	GE healthcare	Mississauga, Ontario, Canada
RNAase	Sigma-Aldrich	Oakville, Ontario, Canada
Sodium dodecyl sulfate (SDS), L3771	Sigma-Aldrich	Oakville, Ontario, Canada
Sodium chloride (NaCl), S671-10	Fisher Scientific	Madison, Wisconsin, USA
T4 DNA LIGASE, M0202S	Biolab	Lawrenceville, Georgia, USA
T4 polynucleotide kinase	NEB	Mississauga, Ontario, Canada
Tris, 0826	AMRESCO	North York, Ontario, Canada
Tris(2-carboxyethyl)phosphine	Sigma-Aldrich	Oakville, Ontario, Canada
Triton™ X-100, X100	Sigma-Aldrich	Oakville, Ontario, Canada
Trypsin-EDTA, T4049	Sigma-Aldrich	Oakville, Ontario, Canada
Trypsin, 85450C	Sigma-Aldrich	Oakville, Ontario, Canada
Tryptone, TRP402.205	Bioshop	Burlington, Ontario, Canada
TWEEN® 20, BP337	Fisher Scientific	Madison, Wisconsin, USA
Q5® Hot Start High-Fidelity DNA Polymerase, M0493S	NEB	Mississauga, Ontario, Canada
Western Blotting detection reagent, RPN2232	GE Health Care	Little Chalfont, UK
Xylene cyanol	Sigma-Aldrich	Oakville, Ontario, Canada

3.2 Plasmids DNA and Mutagenesis

Human DDX43 cDNA clone was purchased from the SPARC BioCentre, the Hospital for Sick Children, Toronto, Canada. DDX43 gene (full-length, N-terminal domain, C-terminal helicase domain, or KH domain) was PCR amplified and cloned into the *Nde*I and *Xho*I sites of a pET28a vector (Novagen). The lysine mutant in motif I K292A, the aspartic acid mutant in DEAD-box in motif II D396A, and KH domain mutants G46D, G84D, and G154D were generated with QuikChange. Site-directed mutagenesis kit (Agilent Technologies) using primers listed in the table **(Table 3.2)**. The KH domain of Sam68 (Lukong and Richard, 2003) was PCR amplified from a pEGFPC1-Sam68 template and cloned into the *Nde*I and *Xho*I sites of a pET28a vector. All plasmids were verified by DNA sequencing.

3.3 RNA and DNA Substrates

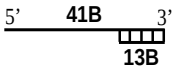
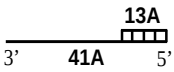
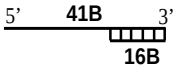
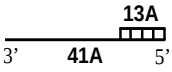
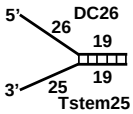
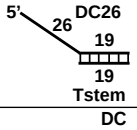
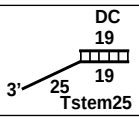
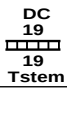
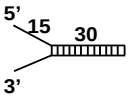
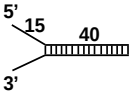
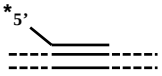
PAGE-purified oligonucleotides were purchased from Integrated DNA Technologies and are listed in (**Table 3.3)**. For each substrate, a single oligonucleotide was 5′-end-labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase (NEB) at 37°C for 1 h. Unincorporated radionucleotides were removed by a G25 chromatography column (GE Healthcare). Single-stranded DNA or RNA substrates were kept at 4°C and ready to use. For the double-stranded DNA substrate, a [γ-^{32}P]ATP-labeled oligonucleotide was annealed to a 2.5-fold excess of the unlabeled complementary strands in annealing buffer (10 mM Tris–HCl, pH 7.5, 50 mM NaCl) by heating at 95°C for 6 min and then cooling slowly to room temperature. For double-stranded RNA, a [γ-^{32}P]ATP-labeled oligonucleotide was annealed to a 2.5-fold excess of the unlabeled complementary strands in annealing buffer (10 mM MOPS, pH 6.5, 1 mM EDTA, 50 mM KCl) by heating at 95°C for 6 min and then cooling slowly to room temperature. All double-stranded substrates were purified by PAGE isolation, and their concentrations were determined by liquid scintillation counting before use.

Table 3.2 List of primers used in the study

Primer	Sequence (5'-3')	Used in this study
DDX43-F-NdeI	ACGT**CATATG**TCCCACCACGGAGGAGCTCCC	Froward primer to PCR amplify full-length DDX43 gene or its N-terminal KH domain for cloning in NdeI site of pET28a vector
DDX43-R-XhoI	GCAT**CTCGAG**ATGAAACTTCTTGGGCCTTCCTTG	Reverse primer to PCR amplify full-length DDX43 gene or its C-terminal helicase domain for cloning in XhoI site of pET28a vector
DDX43-NT-R-XhoI	GCAT**CTCGAG**AACCTCAGGATAACATTGAAAG	Reverse primer to PCR amplify DDX43 gene N-terminal for cloning in XhoI site of pET28a vector
DDX43-HD-F-NdeI	ACGT**CATATG**GAAAACATTAAAAAGGCAGG	Forward primer to PCR amplify DDX43 C-terminal helicase domain for cloning in NdeI site of pET28a vector
DDX43-G46D-F	GAGGGATATAGTGTC**GAC**AGAGGTGGTCGCTGG	Forward primer for site-directed mutagenesis of DDX43-G46D
DDX43-G46D-R	CCAGCGACCACCTCT**GTC**GACACTATATCCCTC	Reverse primer for site-directed mutagenesis of DDX43-G46D
DDX43-G84D-F	GTTGGCGCGGTAATC**GAT**CGTGGTGGGTCAAAAAT	Forward primer for site-directed mutagenesis of DDX43-G84D
DDX43-G84D-R	ATTTTTGACCCACCACG**ATC**GATTACCGCGCCAAC	Reverse primer for site-directed mutagenesis of DDX43-G84D
DDX43-G154D-F	CATTCCAACCTTCTGTT**GAT**AAAGATGGAAGCACAGATAAC	Forward primer for site-directed mutagenesis of DDX43-G154D
DDX43-G154D-R	GTTATCTGTGCTTCCATCTTT**ATC**AACAGAAGGTTGGAATG	Reverse primer for site-directed mutagenesis of DDX43-G154D
DDX43-K292A-F	CAGACTGGAACAGGA**GCG**ACATTGTGTTATTTA	Forward primer for site-directed mutagenesis of DDX43-K292A
DDX43-K292A-R	TAAATAACACAATGT**CGC**TCCTGTTCCAGTCTG	Reverse primer for site-directed mutagenesis of DDX43-K292A
DDX43-D396A-F	CCTACTTGGTTTTA**GCT**GAAGCAGACAAGATG	Forward primer for site-directed mutagenesis of DDX43-D396A
DDX43-D396A-R	CATCTTGTCTGCTTC**AGC**TAAAACCAAGTAGG	Reverse primer for site-directed mutagenesis of DDX43-D396A
Sam68-KH-F-NdeI	ACGT**CATATG**GAGCCAGAGAACAAGTACCTG	Froward primer to PCR amplify KH domain of Sam68 for cloning in NdeI site of pET28a vector
Sam68-KH-R-XhoI	GCAT**CTCGAG**GGGTTCAGGTACTCCATTCAAG	Reverse primer to PCR amplify KH domain of Sam68 for cloning in XhoI site of pET28a vector
DDX43-KH-F-NdeI	ACGT**CATAT**GCCGCTGTGTTTTGCTTTGAAG	Forward primer to PCR amplify DDX43 gene N-terminal KH domain for cloning in NdeI site of pET28a vector
DDX43-KH-F-XhoI	GCA**TCTCG**AGTTCTGAATTGTAATTTTCTT C	Reverse primer to PCR amplify DDX43 gene N-terminal KH domain for cloning in XhoI site of pET28a vector

Restriction digestion sites and mutagenic bases are boldfaced and underlined. To design oligonucleotides ending with G or C, the forward primer and reverse primer may not be in the equal length.

Table 3.3 List of RNA and DNA substrates used in the study

Substrate name	Structure or description	Nucleotide sequence (5'→3')
5' tailed dsRNA (13 bp)		RNA-13B: GCGUCUUUACGGU RNA-41B: AAAACAAAACAAAACAAAACAAAAUAGCACCGUA AAGACGC
3' tailed dsRNA (13 bp)		RNA-13A: ACCGUAAAGACGC RNA-41A: GCGUCUUUACGGUGCUUAAAACAAAACAAAA CAAAACAAAA
Blunt-end dsRNA (13 bp)		RNA-13B: GCGUCUUUACGGU RNA-13A: ACCGUAAAGACGC
5' tailed dsRNA (16 bp)		RNA-16B: GCGUCUUUACGGUGCU RNA-41B: AAAACAAAACAAAACAAAACAAAAUAGCACC GUAAAGACGC
Blunt end DNA:RNA hybrid (13 bp)		DNA-13A: ACCGTAAAGACGC RNA-13B: GCGUCUUUACGGU
3' tailed DNA:RNA hybrid (13 bp)		DNA-41A: GCGTCTTTACGGTGCTTAAAACAAAACAAAA CAAAACAAAA RNA-13A: ACCGUAAAGACGC
Fork dsDNA (19 bp)		DC26: TTTTTTTTTTTTTTTTTTTTTCCCAGTAAAACGA CGGCCAGTGC Tstem25: GCGGTCCCAAAAGGGTCAGTGCTGGCATTTTGCT GCCGGTCACG
5' tailed dsDNA (19 bp)		DC26: TTTTTTTTTTTTTTTTTTTTTCCCAGTAAAACGA CGGCCAGTGC TSTEM:GCACTGGCCGTCGTTTTAC
3' tailed dsDNA (19 bp)		DC: GTAAAACGACGGCCAGTGC Tstem25: GCGGTCCCAAAAGGGTCAGTGCTGGCATTTTGCT GCCGGTCACG
Blunt-end dsDNA (19 bp)		DC: GTAAAACGACGGCCAGTGC TSTEM:GCACTGGCCGTCGTTTTAC
Fork dsDNA (30 bp)		Fork 30/15-T: TTTTTTTTTTTTTTGGTGATGGTGTA TTGAGT GGGATGCATGCA Fork 30/15-B: TGCATGCATCCCACTCAATACACCATCACCTT TTTTTTTTTTTT
Fork dsDNA (40 bp)		Fork 40/15-T: TTTTTTTTTTTTTTTACGAGCTAAATTAGAGC GACTGCACAACTGTAAGGTCCGT Fork 40/15-B: ACGGACCTTACAGTTGTGCAGTCGCTCT AATTTAGCTCGTTTTTTTTTTTTTTT
3'Tail plasmid triplex		The duplex plasmid DNA is prepared above. 3′ tail TC30 is used as third strand.
5'Tail plasmid triplex		The duplex plasmid DNA is prepared above. 5′ tail TC30 is used as third strand.
dT_{30}	Used as ssDNA	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
DNA-30 mer	Used as a ssDNA random sequence	GAGCTACCAGCTACCCCGTATGTCAGAGAG
RNA-30 mer	Used as a ssRNA random sequence	GAGCUACCAGCUACCCCGUAUGUCAGAGAG

3.4 Expression and Purification of Recombinant DDX43 Protein

The plasmid, pET28a-DDX43 was transformed into Rosetta 2 cells (EMD Millipore). Recombinant His-tagged proteins were subjected to a two-step purification using Nickel Affinity beads (Sigma) and a Sephacryl S-300 HR 16/60 gel filtration column (GE Healthcare). The Rosetta 2 cells harboring the recombinant gene were grown at 37°C in LB medium containing 30 μg/mL kanamycin and 34 μg/mL of chloramphenicol until the A600 reached 0.6 and then induced by addition of 0.3 mM IPTG for overnight expression at 15°C. The cells were harvested by centrifugation at 5000 *g* for 10 min at 4°C and stored at -80°C until used. The cell suspension was lysed by sonication in buffer A (25 mM Tris pH 8.0, 0.15 M NaCl, 100 μM Tween 20 and 10% glycerol) having final concentration of 1 mM phenylmethylsulfonyl fluoride (PMSF), protease inhibitor (Roche Applied Science) of at 4°C, with 5 short burst of 10 sec at the intervals of 5 min. The lysed cells were centrifuged at 45,000 g for 30 min at 4°C. The supernatant was applied to the Ni-NTA beads equilibrated with buffer A. The beads were then washed with 10 column volumes (CV) of wash buffer B (25 mM Tris pH 8.0, 0.5 M NaCl, 100 μM Tween 20 and 10% glycerol) containing 25 mM imidazole and the proteins then eluted with 5 CV of elution buffer B containing 250 mM imidazole. The protein fractions were confirmed with SDS-PAGE, and fractions with high protein yield were pooled and subjected to size-exclusion chromatography on a Sephacryl S-300 HR 16/60 equilibrated with buffer A. The protein fractions were collected at a flow rate of 0.5 mL/min with the same buffer. The protein was confirmed with SDS-PAGE and the fractions in the peak were pooled and concentrated. Protein concentration was determined by the Bradford method using bovine serum albumin (BSA) as the standard.

3.5 Western Blot

Proteins (30 μg) were denatured at 100°C for 5 min, then resolved on 10% polyacrylamide Tris-glycine SDS gels, and transferred to PVDF membranes. The membrane was blocked in PBS containing 5% skimmed milk at room temperature for 1 h, followed by probing with a

rabbit polyclonal anti-DDX43 antibody (1:1000, cat# HPA031381, Sigma), or a mouse anti-His-tag monoclonal antibody (1:5000, cat# SAB1400618, Sigma), respectively. Goat anti-rabbit or goat anti-mouse IgG- horseradish peroxidase conjugate (cat#sc-2004, Santa Cruz Biotech) were used as secondary antibody at a 1:10,000 dilution and detected using ECL Plus (GE Healthcare).

3.6 Helicase Unwinding Assays

The helicase assay reaction mixtures (20 μL) contained 40 mM Tris (pH 8.0), 0.5 mM magnesium chloride ($MgCl_2$), 0.15 mM sodium chloride (NaCl), 0.01% Nonidet P-40, 0.1 mM DTT, 1 mg/mL bovine serum albumin (BSA), equimolar mixture of 2 mM ATP and $MgCl_2$, 0.5 nM of the specified duplex RNA and DNA substrate, and the indicated concentrations of DDX43 protein. Helicase reactions were initiated by the addition of DDX43 and incubated at 37°C for 15 min unless otherwise indicated. Reactions were quenched by addition of 20 μL of 2×stop buffer (17.5 mM EDTA, 0.3% SDS, 12.5% glycerol, 0.02% bromphenol blue, 0.02% xylene cyanol). For standard duplex RNA and DNA substrates, a 10-fold excess of the unlabeled oligonucleotide with the same sequence as the labeled strand was included in the quench to prevent reannealing. Reaction products of duplex RNA substrates were resolved on nondenaturing 15% (19:1 acrylamide:bisacrylamide) polyacrylamide gels and products of DNA unwinding reactions were resolved on nondenaturing 12% (19:1 acrylamide:bisacrylamide) polyacrylamide gels. Radiolabeled RNA and DNA species in polyacrylamide gels were visualized using a PhosphorImager and quantitated using Quantity One software (Bio-Rad).

3.7 Translocase Assays

The triplex helicase assay reaction mixtures (20 μL) contained 40 mM Tris (pH 8.0), 0.5 mM magnesium chloride ($MgCl_2$), 0.15 mM sodium chloride (NaCl), 0.01% Nonidet P-40, 0.1 mM DTT, 1 mg/mL bovine serum albumin (BSA), equimolar mixture of 2 mM ATP and

$MgCl_2$, 0.5 nM of the specified 5'-tail and 3'-tail plasmid-triplex substrate, and the indicated concentrations of DDX43 protein. Helicase reactions were initiated by the addition of DDX43 and incubated at RT for 30 min unless otherwise indicated. Reactions were quenched by addition of 10 µL of 2×stop buffer (17.5 mM EDTA, 0.3% SDS, 12.5% glycerol, 0.02% bromphenol blue, 0.02% xylene cyanol). The products of the helicase reactions for triplex DNA substrates were resolved on 10% (19:1 acrylamide:bisacrylamide) polyacrylamide gels with 40 mM Tris acetate (pH 5.5) and 25% glycerol and running buffer containing 40 mM Tris acetate, pH 5.5, and 5 mM $MgCl_2$. Radiolabeled species in polyacrylamide gels were visualized using a PhosphorImager and quantitated using Quantity One software (Bio-Rad).

3.8 Annealing Assay

The reaction (20 µL) was carried out with 0.5 nM of radiolabeled 19 bp forked dsDNA. The substrate was denatured first at 100°C for 5 minutes and then incubated with 0 to 3 µM of DDX43 protein at 37^0C for 15 min with or without 2 mM ATP. After incubation the reaction was stopped by addition of equal volume of 2×stop buffer (17.5 mm EDTA, 0.3% SDS, 12.5% glycerol, 0.02% bromophenol blue, 0.02% xylene cyanol). The mixture was resolved on 12% native PAGE gel for 2 h at 180 V. The resolved radiolabeled species were visualized using Phosphor-Imager (Bio-Rad) and analyzed using Quantity One software (Bio-Rad).

3.9 Strand Exchange Assay

The Strand exchange assay was performed in the same manner as the helicase unwinding assay, except cold oligo was included in the reaction. The reaction mixture (20 µL) for the assay contained 40 mM Tris (pH 8.0), 0.5 mM magnesium chloride ($MgCl_2$), 0.15 mM sodium chloride (NaCl), 0.01% Nonidet P-40, 0.1 mM DTT, 1 mg/mL bovine serum albumin (BSA), equimolar mixture of 2 mM ATP and $MgCl_2$, 0.5 nM of the specified duplex DNA substrate. The reaction was initiated by the addition of DDX43 and incubated at 37°C for 15 min unless otherwise indicated. Reactions were quenched by addition of 20 µL of 2×stop

buffer (17.5 mM EDTA, 0.3% SDS, 12.5% glycerol, 0.02% bromphenol blue, 0.02% xylene cyanol). Products of DNA strand exchange activity were resolved on nondenaturing 12% (19:1 acrylamide:bisacrylamide) polyacrylamide gels.

3.10 Electrophoretic Mobility Shift Assay (EMSA)

Protein/DNA or RNA binding mixtures (20 µL) contained the indicated concentrations of DDX43 and 0.5 nM of the specified ^{32}P-end-labeled DNA substrate in the same reaction buffer as that used for helicase assays (see above) without ATP. The binding mixtures were incubated at room temperature for 30 min after the addition of DDX43. After incubation, 3 µL of loading dye (74% glycerol, 0.01% xylene cyanol, 0.01% bromphenol blue) was added to each mixture, and samples were loaded onto native 5% (19:1 acrylamide/bisacrylamide) polyacrylamide gels and electrophoresed at 200 V for 2 h at 4°C using 1×TBE as the running buffer. The resolved radiolabeled species were visualized using a PhosphorImager and analyzed with Quantity One software (Bio-Rad).

3.11 ATP Hydrolysis Assays

ATP hydrolysis was measured using [γ-^{32}P] ATP (PerkinElmer Life Sciences) and analyzed by thin layer chromatography (TLC) on polyethyleneimine-cellulose plates (J.T. Baker). The standard reaction mixture (50 µL total volume) containing 40 mM Tris (pH 8.0), 0.5 mM $MgCl_2$, 0 15 mM NaCl, 0.01% Nonidet P-40, 0.1 mM DTT, 1 mg/mL BSA, 250 µM [γ-^{32}P] ATP, 2.5 µM DDX43 protein, ssDNA dT30 or ssRNA rU_{30} stimulator (0 to 10 mM), was incubated at 37°C for 30 min. Reactions were quenched with 50 mM EDTA final concentration. The reaction mixture was spotted onto a PEI-cellulose TLC plate and resolved by using 0.5 M LiCl, 1 M formic acid as the carrier solvent. The TLC plate was exposed to a phosphorimager cassette for 30 min and visualized using a Phosphor-Imager and analyzed with Quantity One software (Bio-Rad).

3.12 Cell Culture

The cells were cultured in 10-cm petri dishes under conditions of 5% CO2 and 37°C. Unless otherwise stated, all cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM) media (Cat# 12-604F, Lonza), supplemented with 10% fetal bovine serum (Cat# F1051, Sigma), 50 µg/mL penicillin streptomycin (Cat# P4333, Sigma). For subculture, cells were detached from the plate floor by trypsinization. Cells were subcultured 2-3 times per week at a dilution of 1:5 - 1:20, depending on the cell line. The working area was decontaminated with 70% ethanol before and after use and cell culture work was performed under sterile conditions.

3.13 Immunoprecipitation

HEK293T cells cultured in 10-cm plates were transfected with DNA (pcDNA-3XFLAG-DDX43) using 1% polyethyleneimine 'Max' (PEI) at a ratio of 1:4. For each plate, 10 µg of the appropriate DNA was mixed with 500 µL of 0.15 M sterile NaCl via gentle vortex for 10 sec. Then 40 µL of the transfection reagent, PEI, was added to this mixture followed by another 10 sec of a gentle vortex. DNA–PEI complex was formed by incubating the mixture at room temperature for 10 min followed by dispensing the complex dropwise into the plates. The cells were incubated for 24 h post transfection before harvest. For immunoprecipitation, cells were washed in cold PBS, and lysed in freshly prepared lysis buffer constituting 20 mM Tris (pH 8.0), 0.5 % Nonidet P-40, 100 mM NaCl, 2 mM EDTA, 1 mM PMSF and protease inhibitors 0.5 µg/mL of DNAase and RNAase for 30 min followed by centrifugation for 30 min at 40,000 *g* at 4°C. Supernatants were collected into fresh tubes and incubated with anti-FLAG resin (Sigma) and maintained on a rotator for 1 h at 4°C. The beads were washed with lysis buffer for 3 times and incubated with elution buffer constituting 25 mM Tris (pH 7.4), 0.01 % Nonidet P-40, 150 mM NaCl, 1 mM EDTA, 10% glycerol, 1 mM DTT, 1 mM TCEP, protease inhibitor, and 150 µg/mL 3XFLAG peptide. The immunoprecipitated proteins were resolved on 10% gradient SDS-PAGE (Bio-Rad) and visualized by silver staining. The

identities of excised bands were determined by the mass spectrometry service at the University of Alberta.

4. RESULTS

4.1 Protein Overexpression and Purification

4.1.1 DDX43 Protein Exists as a Monomer in Solution

To characterize the DDX43 protein, we cloned the human *DDX43* gene into a pET28a vector, overexpressed it in *E. coli*. and purified it by two-step chromatography. First, the His-tagged DDX43 protein was passed through a Ni-NTA affinity chromatography, and the retained proteins were eluted with imidazole. The recombinant DDX43 proteins were purified to near homogeneity as judged by their appearance on coomassie-stained SDS-polyacrylamide gel **(Fig. 4.1 A).** The fractions with high protein concentration were pooled and applied to a Sephacryl S-300 size exclusion column. DDX43 protein eluted in two major peaks, namely peak 1 and peak 2 (**Fig. 4.1 B**). Identical migration bands were observed on the coomassie stained SDS-PAGE gel for these two peaks (**Fig. 4.1 C**). The identity of the DDX43 protein was confirmed by Western blot using anti-DDX43 and anti-His antibodies respectively (**Fig. 4.1 D**), indicating both fractions are indeed DDX43 protein. According to the molecular mass standards used to calibrate the size exclusion column, peak 1 was in the void volume of the column, likely representing aggregated DDX43 protein, while the mass of peak 2 was about 67.4 ± 5.1 kDa, which is close to the predicted monomeric form of DDX43 (72.4 kDa). Therefore, we collected the monomer fractions of DDX43 for the following biochemical assays.

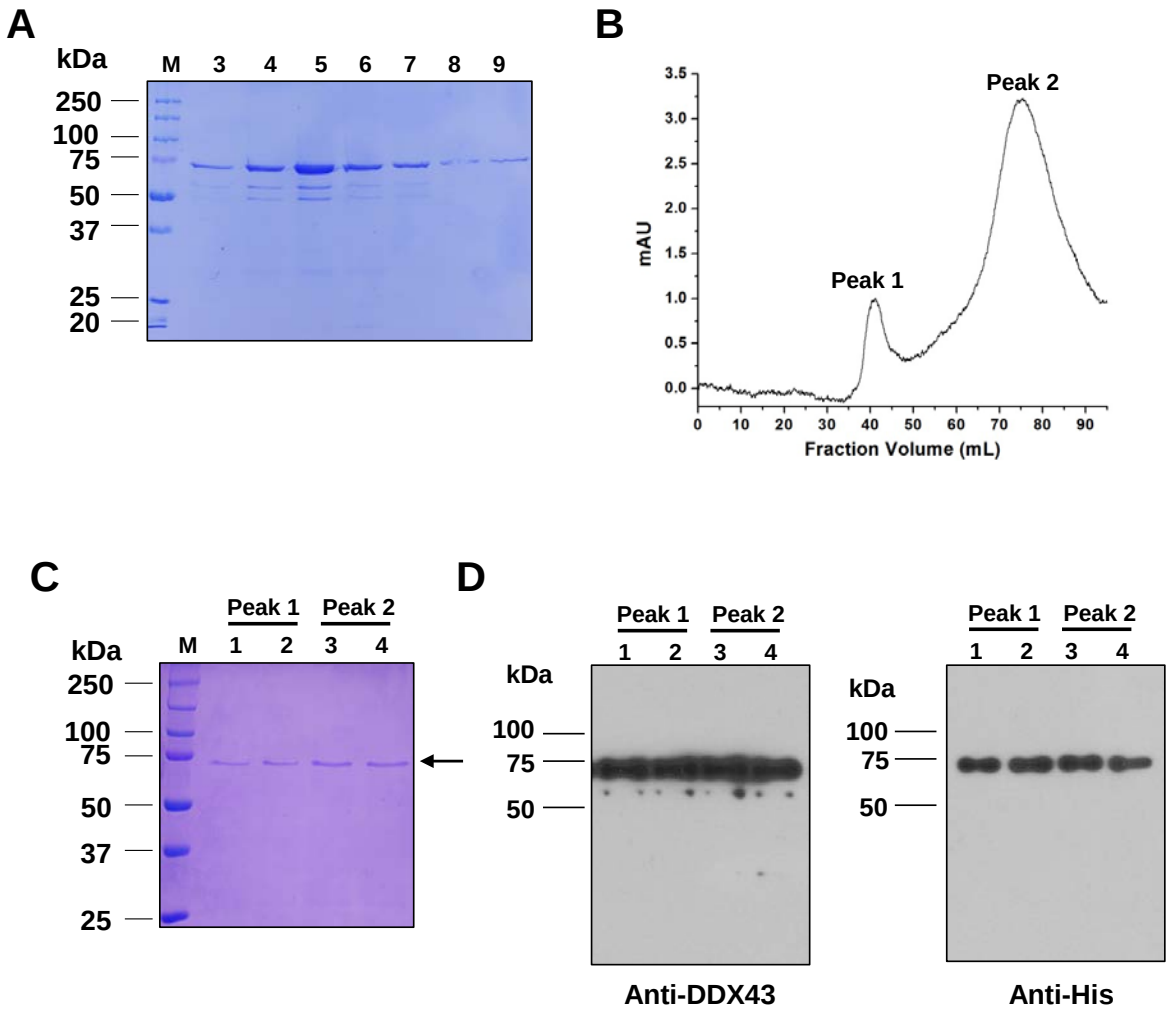


Figure 4.1 Purification and identification of DDX43 protein. (**A**) SDS-PAGE analysis of the eluted DDX43 fractions from a Ni-NTA column. M, marker. Fractions 3-9 are shown. (**B**) Chromatographic profiles of recombinant DDX43 proteins eluting from a Sephacryl S-300 HR column. Two peaks are indicated. (**C**) SDS-PAGE analysis of the peaks shown in B. (**D**) Western blot analyses of the proteins shown in C with antibody against DDX43 (left) or His (right).

4.2 DDX43 Protein Unwinds RNA Substrates

4.2.1 DDX43 Protein Preferably Unwinds RNA in a 5' to 3' Direction

According to its primary sequence, DDX43 belongs to the DEAD-box RNA helicase family, thus, I started to characterize its RNA unwinding activity. Using a 5'tail 13-bp duplex RNA,

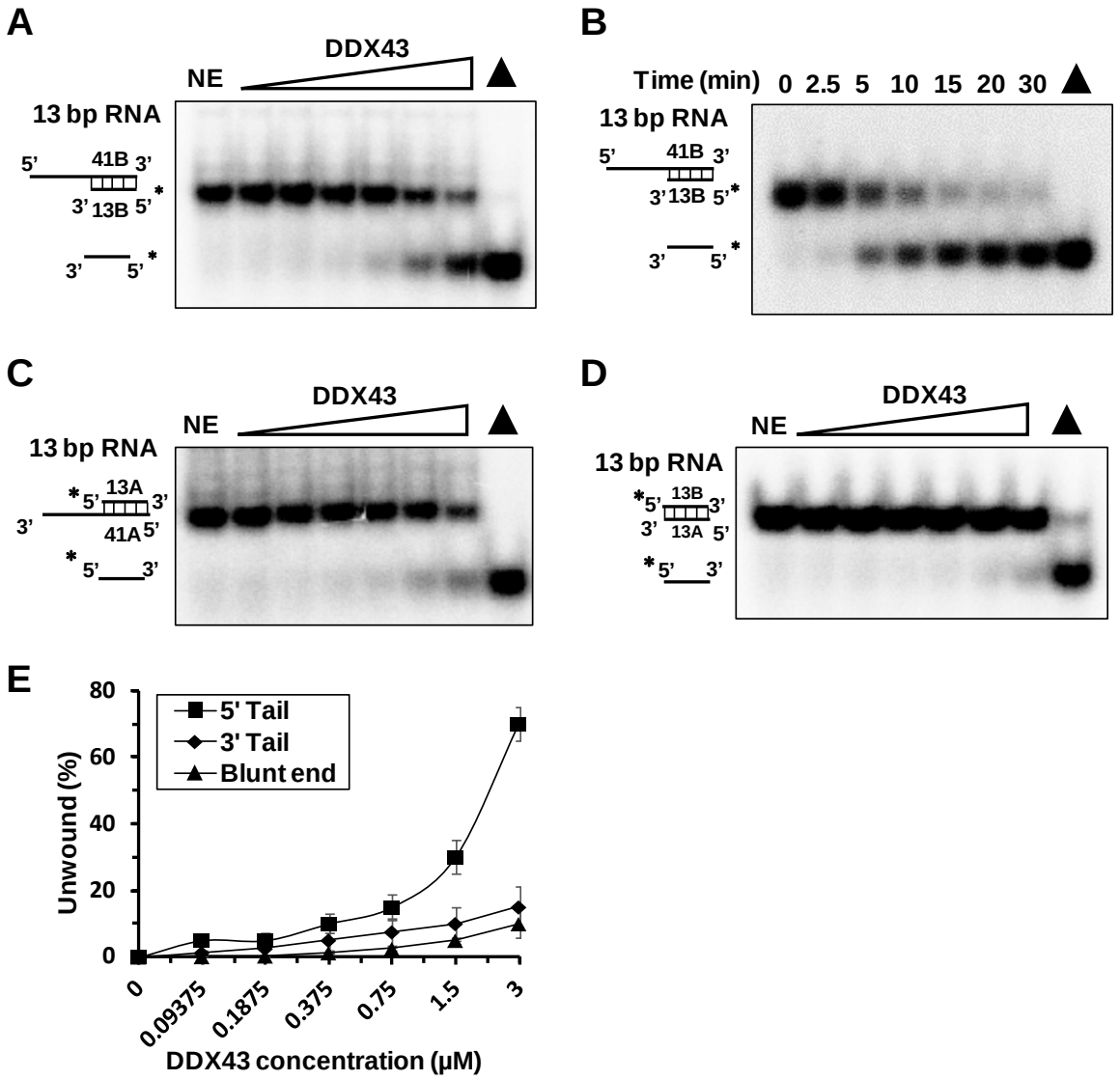


Figure 4.2 RNA helicase assays of recombinant DDX43 proteins. (**A**) A representative image of helicase reactions performed by incubating 0.5 nM 5' tail 13 bp duplex RNA substrate with increasing protein concentration (0.09, 0.18, 0.37, 0.75, 1.5, 3 μM) at 37 °C for 15 min. (**B**) A representative image of helicase reactions performed by incubating 0.5 nM 5' tail 13 bp duplex RNA substrate and 150 nM DDX43 protein with increasing time (0-30 min) at 37 °C. (**C, D**) Representative images of helicase reactions performed by incubating 0.5 nM 3' tail 13-bp duplex RNA substrate (**C**) and blunt-end 13 bp duplex RNA substrate (**D**) with increasing protein concentration (0.09, 0.18, 0.37, 0.75, 1.5, 3 μM) at 37 °C for 15 min. (**E**) Quantitative analyses of RNA unwinding of DDX43 in panel A, C and D. Data represents the mean of at least three independent experiments with SD indicated by error bars. NE, no enzyme; WT, wild-type; filled triangle, heat denatured substrate control.

I found that DDX43 could efficiently unwind the substrate in the presence of ATP in both a concentration (0-3 µM) and time-dependent manner (0-30 min) **(Fig. 4.2 A and B)**. However, it exhibited poor unwinding activity on both a 3' tail 13-bp duplex RNA **(Fig. 4.2 C)** and blunt end RNA substrate **(Fig. 4.2 D)**. These results indicated that DDX43 unwinds RNA substrates irrespective of the single strand overhang but with a strong preference for 5' tail RNA substrate.

4.2.2 DDX43 Protein Exhibits Low Processivity on RNA Substrates

In general, DEAD-box helicases have been found to have low processivity on the RNA substrate (Jankowsky and Fairman, 2007). Thus to examine the processivity of DDX43 on RNA substrates, I increased the length for the duplex RNA substrate from 13 bp to 16 bp under the same reaction conditions and observed that DDX43 almost failed to unwind the longer substrate. These results indicated that DDX43 has low processivity on longer RNA substrates (**Fig. 4.3**).

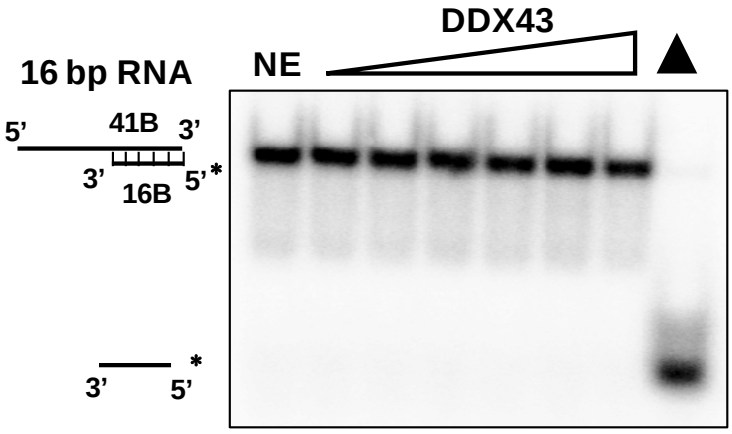


Figure 4.3. Helicase assay for processivity of recombinant DDX43 proteins on an RNA substrate. A representative image of helicase reactions performed by incubating 0.5 nM 5' tail 16 bp duplex RNA substrate with increasing protein concentration (0.09, 0.18, 0.37, 0.75, 1.5, 3 µM) at 37 °C for 15 min. Filled triangle, heat denatured substrate control.

4.3 DDX43 Protein Unwinds DNA Substrates

4.3.1 DDX43 Protein Unwinds DNA in only 3' to 5' Direction

Many helicases have been found to possess dual unwinding activity using both DNA and RNA substrates (Kawaoka and Pyle, 2005). Thus next we asked whether DDX43 can work on DNA substrates.

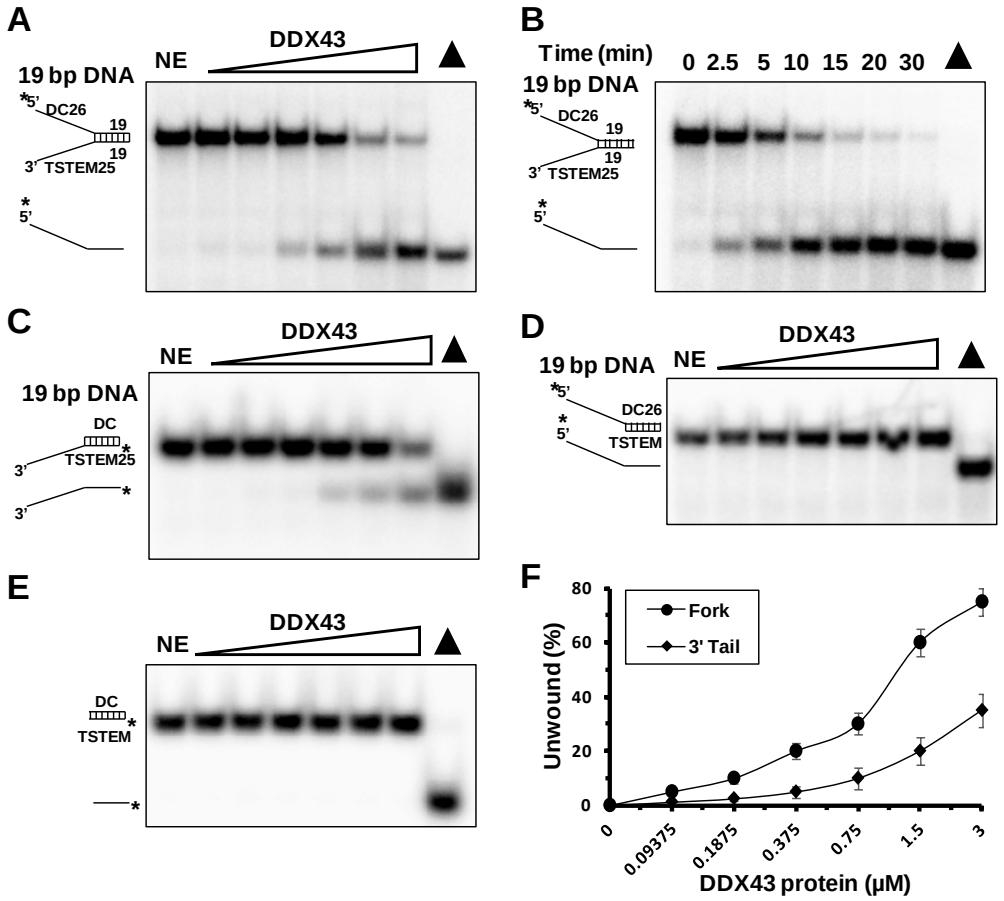


Figure 4.4 DNA helicase assays of recombinant DDX43 proteins. (**A**) A representative image of helicase reactions performed by incubating 0.5 nM 19-bp forked duplex DNA substrate with increasing protein concentration (0.09, 0.18, 0.37, 0.75, 1.5, 3 μM) at 37 °C for 15 min. (**B**) A representative image of helicase reactions performed by incubating 0.5 nM 19-bp forked duplex DNA substrate and 150 nM DDX43 protein with increasing time (0-30 min) at 37 °C. (**C, D**) Representative images of helicase reactions performed by incubating 0.5 nM 3' tail 19-bp duplex DNA substrate (**C**) or 5' tail 19-bp forked duplex DNA substrate (**D**) or blunt-end 19-bp duplex DNA substrate (**E**) with increasing protein concentration (0.09, 0.18, 0.37, 0.75, 1.5, 3 μM) at 37 °C for 15 min. Filled triangle, heat denatured substrate control. (**F**) Quantitative analyses of DNA unwinding of DDX43 in panel A and C Data represents the mean of at least three independent experiments with SD indicated by error bars.

Using a 19-bp forked duplex DNA substrate, we detected that DDX43 can efficiently unwind the DNA substrate in the presence of ATP in both a concentration (0 – 3 μM) and time-dependent manner (0 - 30 min) **(Fig. 4.4 A and B)**. Furthermore, we found that DDX43 was active on a 3’ tail 19-bp duplex DNA substrate, but not a 5’ tail or blunt-end substrate **(Fig. 4.4 C-E)**, indicating that DDX43 translocates in the 3’-5’ direction on a DNA substrate.

4.3.2 DDX43 Protein Exhibits High Processivity on DNA Substrates

Although DDX43 possess low processivity on the RNA substrates, processivity of helicases are important for helicase function in DNA replication and other cellular functions. Thus we aimed to examine the processivity of DDX43 on DNA substrates and we found that DDX43 had very high processivity on DNA substrate as it could efficiently unwind 30 bp and 40 bp forked duplex substrates in a similar fashion **(Fig. 4.5 A and B).**

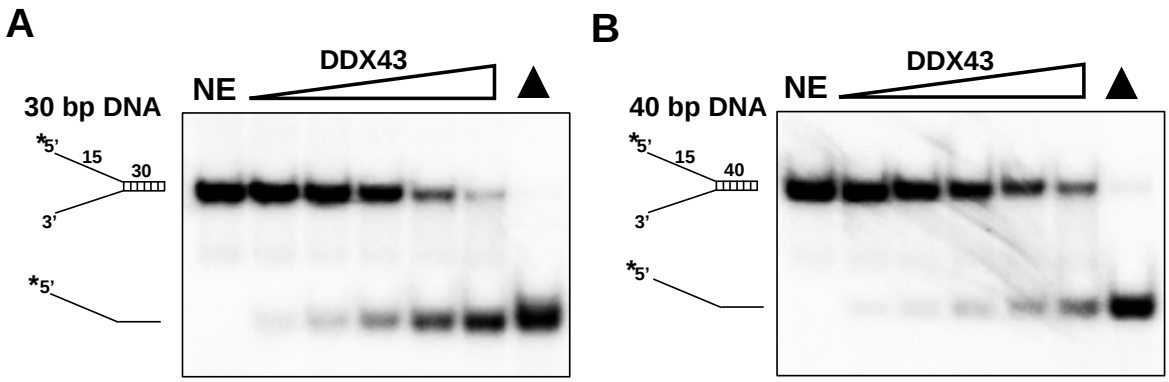


Figure 4.5 Helicase assays for processivity of recombinant DDX43 proteins on DNA substrate. **(A, B)** Representative images of helicase reactions performed by incubating 0.5 nM 30-bp forked duplex DNA substrate (**A**) 40-bp forked duplex DNA substrate (**B**) with increasing protein concentration (0.09, 0.18, 0.37, 0.75, 1.5, 3 μM) at 37 °C for 15 min. Filled triangle, heat denatured substrate control.

4.4 DDX43 has Weak 3’-5’ Translocase Activity

DDX43 melts dsDNA from 3’ to 5’, but not from 5’ to 3’, indicating it might be a 3' to 5' translocase. Triplex displacement experiments have been utilized to monitor the translocase

activity of helicases, including AddAB (Gilhooly and Dillingham, 2014), FANCJ (Meetei *et al*., 2005; Sommers *et al*., 2009) and ChlR1 (Guo *et al*., 2015). In this assay, a triple helix is formed when a third strand forms Hoogsteen base pairs with duplex DNA. If a translocase proceeds through the triplex, it will displace the third strand. Indeed, DDX43 could not displace a 5' tailed triplex, (**Fig. 4.6 A**); but it could displace a 3' tailed triplex in 3' to 5' direction, which indicates that DDX43 possess translocase activity. (**Fig. 4.6 B).**

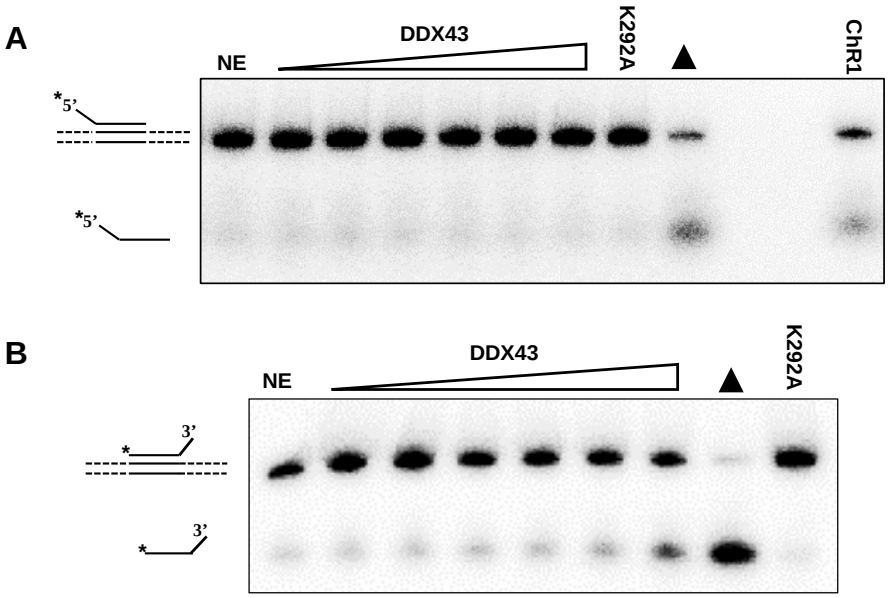


Figure 4.6 Triplex assays of recombinant DDX43 proteins. (**A-B**) A representative image of a triplex assay performed by incubating 5' tail plasmid-triplex **(A)** and 3' tail plasmid-triplex substrate with increasing protein concentration (0.09, 0.18, 0.37, 0.75, 1.5, 3 µM) at RT for 30 minutes **(B)**. ChlR1 was used as a positive control. Filled triangle, heat denatured substrate control.

4.5 DDX43-K292A and D396A Mutants Abolish Helicase Unwinding Activity

To confirm that the unwinding activity detected was truly dependent on DDX43 and not because of contaminants in the preparation of the DDX43 protein, we changed the conserved lysine to alanine (K292A) in motif I, which is essential for ATP binding and hydrolysis, and aspartic acid to alanine in motif II of the DEAD-box (D396A), which helps to bind cation for enzyme activity **(Fig. 1.5 A).** Under identical methods and conditions as

wild-type DDX43, the mutant proteins were purified to near homogeneity (**Fig. 4.7 A**). Helicase assays revealed that neither K292A nor D396A displayed any unwinding activity on the 5' tailed 13 bp duplex RNA (Fig. 4.7 B and C) and 19 bp forked DNA substrate (Fig. 4.7 D and E).

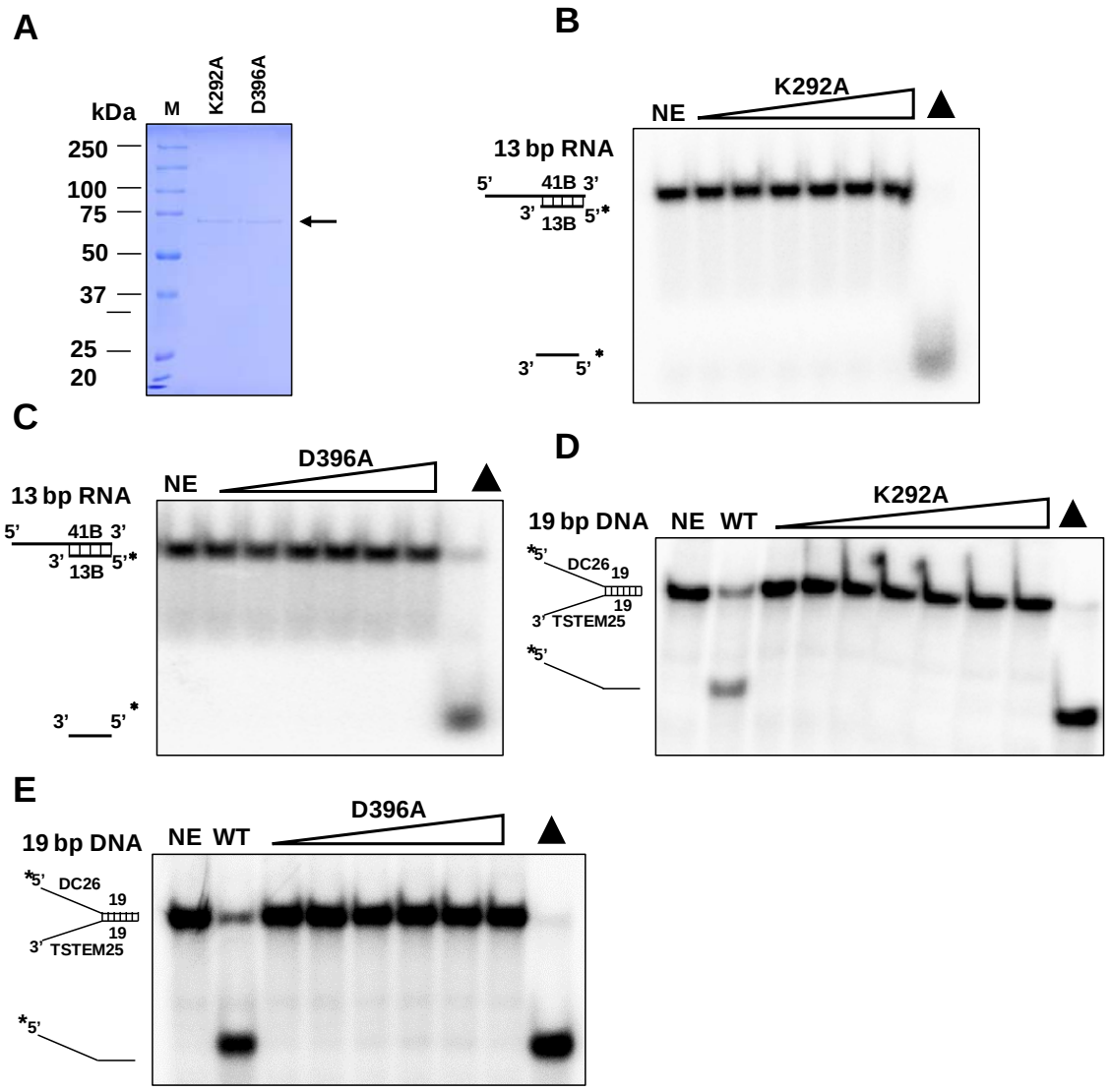


Figure 4.7 Purification and characterization of mutants K292A and D396A. (**A**) SDS-PAGE analysis of DDX43 K292A and D396A proteins eluting from a Sephacryl S-300 HR column. (**B, C**) Representative images of helicase reactions performed by incubating 0.5 nM 5 tail 13 bp duplex RNA substrate with increasing protein concentration of K292A **(B)** or D396A **(C)** at 37 °C for 15 min. (**D, E**) Representative images of helicase reactions performed by incubating 0.5 nM 19 bp forked dsDNA substrate with increasing protein concentration of K292A **(D)** or D396A **(E)** at 37 °C for 15 min. Filled triangle, heat denatured substrate control.

4.6 DDX43 Protein Unwinds DNA:RNA hybrid Substrates

Some DEAD-box helicases have been found to unwind DNA-RNA chimeric substrates as well, e.g., Ded1 (Yang and Jankowsky, 2006) and DDX1 (Li *et al*., 2008). Thus we asked whether DDX43 could unwind DNA:RNA hybrid substrates. Using a blunt end 13bp DNA:RNA hybrid substrate, we detected that DDX43 protein could unwind a substrate in a concentration (0 – 3 μM) dependent manner in the presence of the ATP **(Fig. 4.8).**

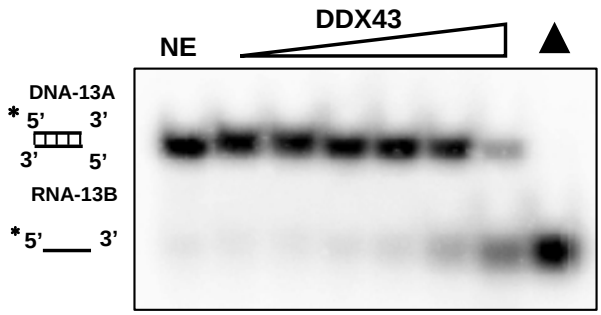


Figure 4.8 Unwinding activity of DDX43 protein on a DNA:RNA hybrid. A representative image of helicase reactions performed by incubating 0.5 nM 5' end labelled 13 bp DNA:RNA blunt-end hybrid substrate with increasing protein concentration (0.09, 0.18, 0.37, 0.75, 1.5, 3 μM) at 37 °C for 15 min. Filled triangle, heat denatured substrate control.

4.7 ATP Hydrolysis and Mg^{2+} are Essential for the Unwinding Activity of DDX43 Protein

Helicases are motor proteins that couple conformational changes induced by ATP binding and hydrolysis with unwinding of duplex nucleic acids (von Hippel and Delagoutte, 2001). To check whether ATP analogs can cause this conformational change in DDX43, I used non-hydrolyzable ATP-analogs, ATP γ S and AMP-PNP and some other ATP analogs to detect the unwinding activity.

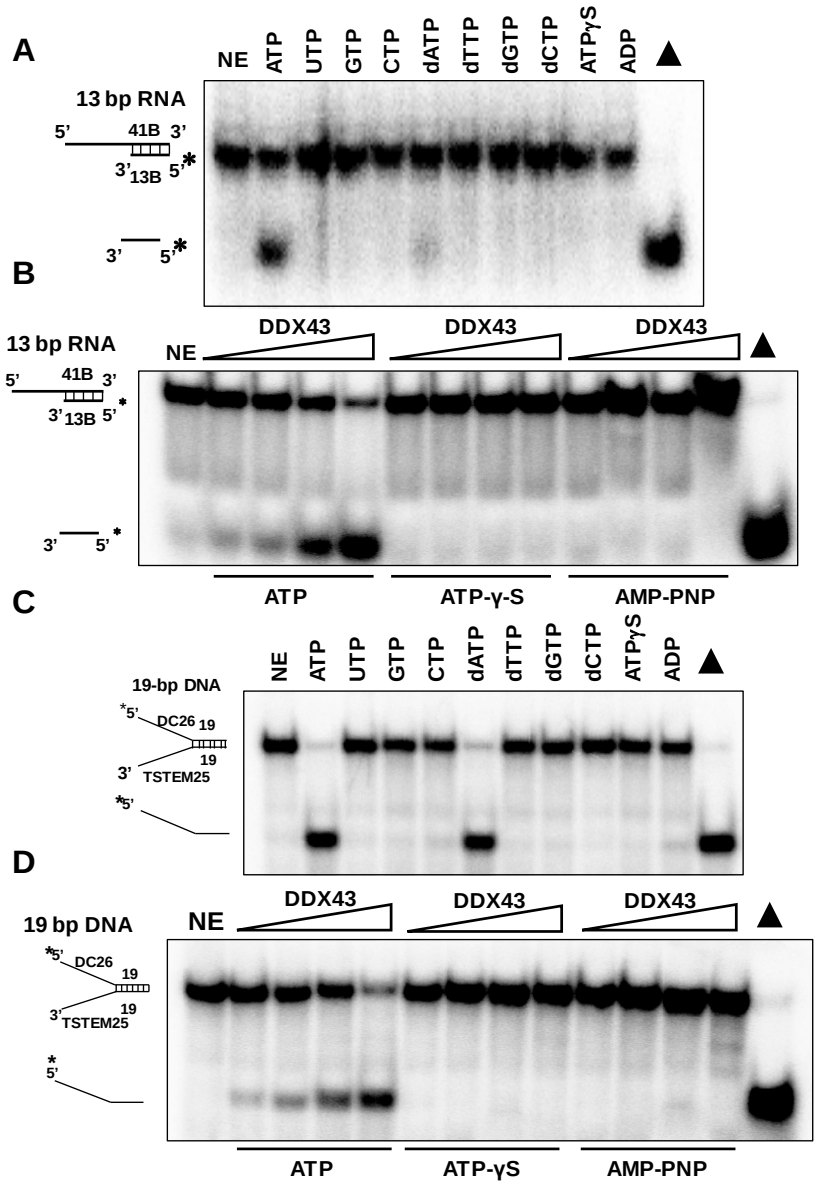


Figure 4.9 RNA and DNA helicase activity of recombinant DDX43 protein in presence of different NTP's and their analogues. (A-B) A representative images of helicase reactions performed by incubating 0.5 nM 5' tail 13 bp duplex RNA substrate with 150 nM DDX43 protein at 37 °C for 15 min with different nucleoside-triphosphate (**A)** or ATP or ATP analogs (**B**). **(C-D)** A representative image of helicase reactions performed by incubating 0.5 nM 5' tail 19 bp forked DNA Substrate with 150 nM DDX43 protein at 37 °C for 15 min with different nucleoside-triphosphate (**C)** or ATP or ATP analogs (**D**). Filled triangle, heat denatured substrate control.

Using a 5' tail 13-bp duplex RNA **(Fig. 4.9 A and B)** and 19-bp forked dsDNA **(Fig. 4.9 C and D)**, I found that DDX43 (0-3 µM) could efficiently unwind both substrates in the

presence of ATP but it did not exhibit any unwinding activity with ATP γ S or AMP-PNP, indicating ATP hydrolysis is essential for the unwinding activity of DDX43.Along with the ATP, the presence of a divalent cation in the helicase core of the enzyme plays a very important role in its unwinding activity (Frick *et al*., 2007).

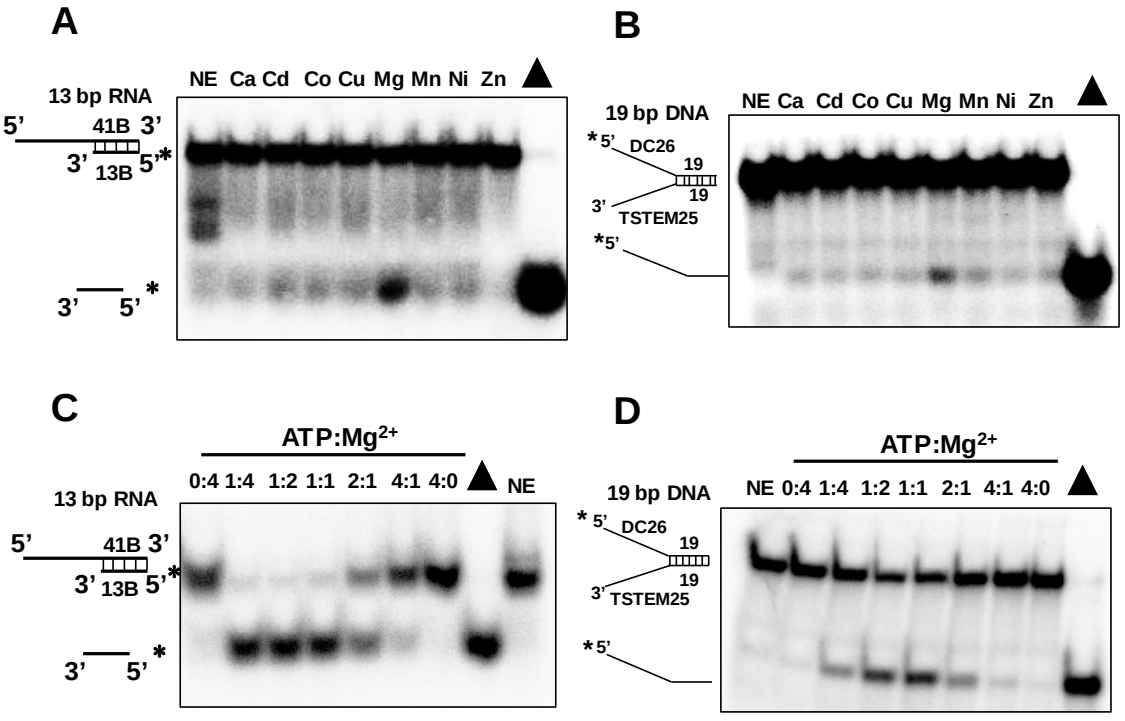


Figure 4.10 Optimized cation and ATP for DDX43 helicase activity. (A, B) Representative images of helicase reactions performed by incubating 0.5 nM 5' tailed 13 bp duplex RNA (**A**) or 19 bp forked duplex DNA substrate (**B**) with indicated cations (2 mM) at the protein concentration of 3 µM at 37°C for 15 min. (**C, D**) Representative images of helicase reactions performed by incubating 0.5 nM 5' tailed 13 bp duplex RNA (**C**) or 19 bp forked duplex DNA substrate (**D**) with indicated ATP:Mg^{2+} ratio at the protein concentration of 3 µM at 37°C for 15 min. Filled triangle, heat denatured substrate control.

To determine which cation is the best co-factor for DDX43, we performed helicase assays with both duplex RNA and DNA in the presence of different divalent cations. Out of eight cations tested (all at 2 mM), Mg^{2+} was found to be the best cation for the helicase activity of DDX43 (**Fig. 4.10 A and B**). Further, we wanted to determine the optimal ATP: Mg^{2+} ratio for the unwinding activity of DDX43. As shown in **(Fig. 4.10 C and D),** ATP: Mg^{2+}

ratios of 1:1 and 1:2 were best for DDX43 unwinding activity.

4.8 DDX43 Protein Does not Possess Annealing Activity

A few DEAD-box RNA helicases promote not only RNA unwinding but also strand annealing. For example, nucleolar RNA helicase 2 (RH2; also known as Gu) and DDX21 shows annealing activity. (Uhlmann-Schiffler *et al*., 2006; Yang and Jankowsky, 2005; Halls *et al*., 2007).

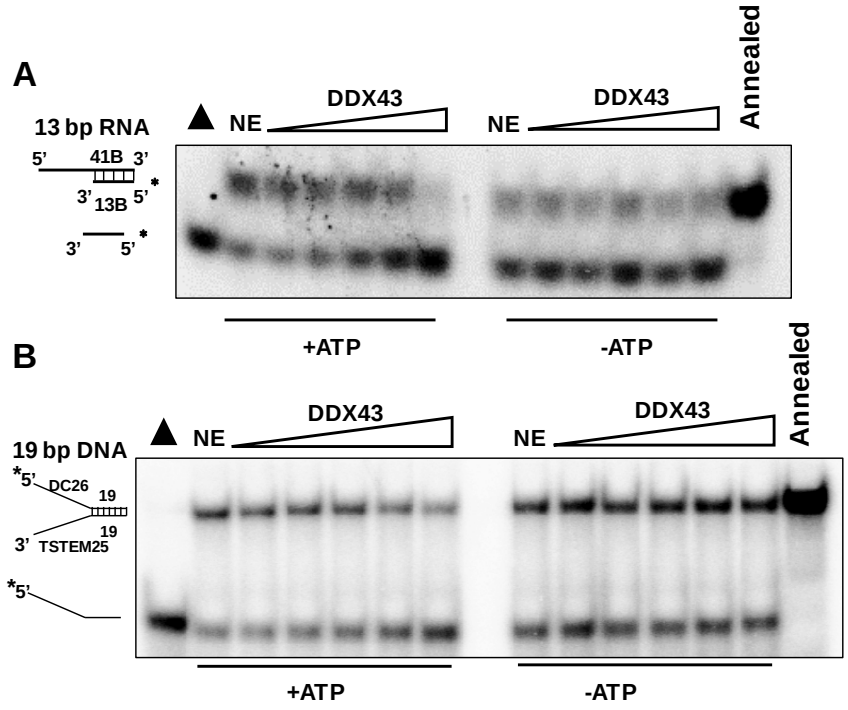


Figure 4.11 Annealing activity of DDX43 protein. (**A**-**B**) A representative image of annealing reactions performed by incubating 0.5 nM 5' tail 13 bp duplex RNA substrate **(A)** or 19 bp forked duplex DNA substrate **(B)** with increasing protein concentration (0.09, 0.18, 0.37, 0.75, 1.5, 3 µM) at 37 °C for 15 min, with and without ATP. Filled triangle, heat denatured substrate control.

I evaluated the annealing activity of the DDX43 using 5' tail 13 bp duplex RNA and 19 bp forked dsDNA in concentration-dependent manner (0 to 3 µM), with and without ATP. At first, the respective substrates were denatured at 100°C for 5 min and then incubated with 0 to 3 µM of DDX43 protein at 37°C for 15 min. Our results indicated that DDX43 does not possess annealing activity on both 13 bp duplex RNA **(Fig. 4.11 A)** or 19 bp forked duplex

DNA substrate (**Fig. 4.11 B**).

4.9 Role of the N-terminal Accessory domain of DDX43 Protein

4.9.1 The KH domain in the N-terminus is involved in Nucleic Acid Binding

From its primary sequence, DDX43 contains three potential KH domains that have the signature GXXG motif (**Fig. 1.5 B**). Also because a KH domain is approximately ~70 amino acids long, we cloned the KH domains-containing the N-terminal region (1-253 aa, named $DDX43^{NT}$) into a pET28 vector for protein production. Previous studies have shown that the first glycine residue in the GXXG motif is essential for the RNA binding function of various KH domains (Barnes *et al*., 2015; Lewis *et al*., 2000; Zhou *et al*., 2002), thus, we changed the glycine to aspartic acid in each motif, namely G46D, G84D, and G154D. Three mutant proteins were purified along with wildtype to near homogeneity (**Fig. 4.12 A**). Using the 19 bp forked duplex DNA substrate, EMSA revealed that G46D and G154D bound DNA similar to or even tighter than wild-type protein (**Fig. 4.12 B**); however, the G84D mutation ($DDX43^{NT(G84D)}$) abolished its binding activity. Similar results were obtained with an ssRNA substrate and ssDNA (**Fig. 4.12 C and D**). However, none of them bound blunt-end dsDNA (**Fig. 4.12 E**).

On the other hand, close inspection of the second KH motif revealed that it has an additional conserved KH domain sequence VIGXXGXXI, shared with some well-known KH domain-containing proteins (**Fig. 4.13 A**); however, the first and third GXXG motifs did not have these additional conserved sequences. Moreover, these amino acids are conserved in DDX43 orthologues across species as well (**Fig. 4.13 B**). To experimentally confirm that the second GXXG motif containing region is a real KH domain, we cloned this region (69-142 aa, named $DDX43^{KH}$) into a pET28 vector, and purified this protein to near homogeneity (**Fig. 4.14 A**). Using the 19 bp forked duplex DNA substrate, EMSA revealed

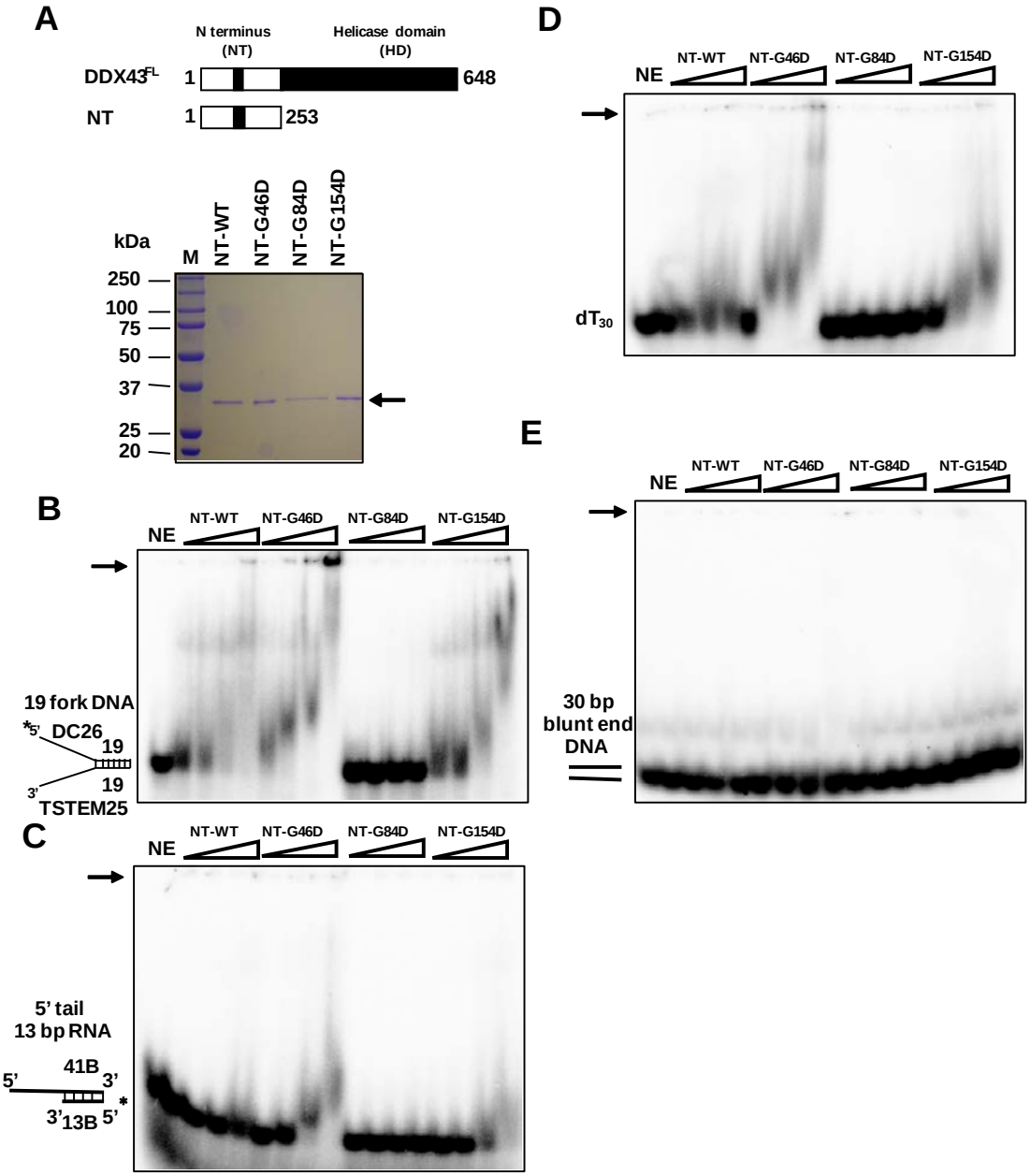
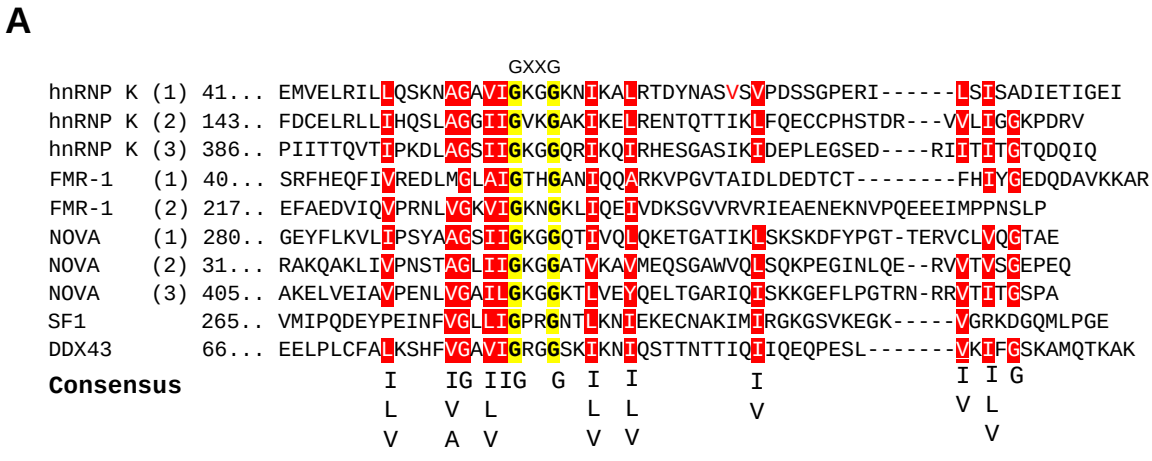


Figure 4.12 Purification and characterization of the DDX43 N-terminal domain. **(A)** Schematic representation of full length DDX43 and DDX43 N-terminal (top), and purified proteins (N-terminal wild type and mutants, 1 µg each) shown on Coomassie stained SDS-PAGE gel (bottom). **(B-E)** Representative EMSA images for binding of potential different KH domains in the N-terminal of DDX43 with 0.5 nM 19 bp forked duplex DNA **(B)** 5' tail 13 bp duplex RNA **(C)** dT_{30} ssDNA **(D)** or 30 bp blunt end dsDNA **(E).** The proteins at various concentrations (0.09, 0.18, 0.37, 0.75, 1.5, 3 µM) were incubated with DNA or RNA substrates, and nucleic acids-protein complexes were resolved on native 5% polyacrylamide gels.

that DDX43^{KH} bound DNA similar to or even tighter than wild-type protein (**Fig. 4.14 B**); however, the G84D mutation in the KH domain (DDX43$^{KH(G84D)}$) abolished its binding activity. Similar results were obtained with a 5' tailed RNA substrate as well as ssDNA and ssRNA (**Fig. 4.14 C, D and E**). However, none of them bound blunt-end dsDNA **(Fig. 4.14 F)**. Collectively, we concluded that the second GXXG motif-containing region is a functional KH domain and is involved in nucleic acid binding in DDX43.



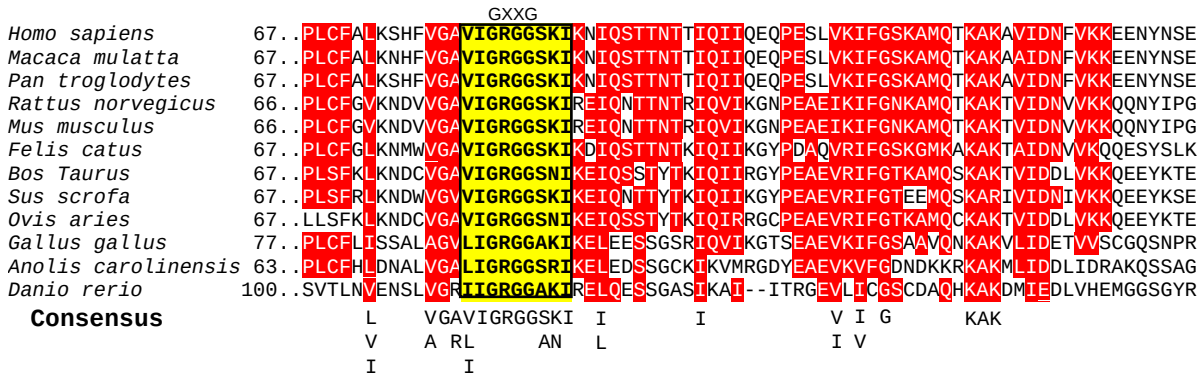


Figure 4.13 KH domain alignment. (A) KH domain sequences of human hnRNP K, FMR1, NOVA and SF1 are aligned with the KH domain sequence of DDX43. **(B)** KH domain alignment of DDX43 orthologous species. The KH domain signature sequence in the middle of the domain is outlined. Numbers on the left indicate positions of the initial amino acid of the KH domain. Highly conserved positions are highlighted, and gaps are introduced for optimal alignment.

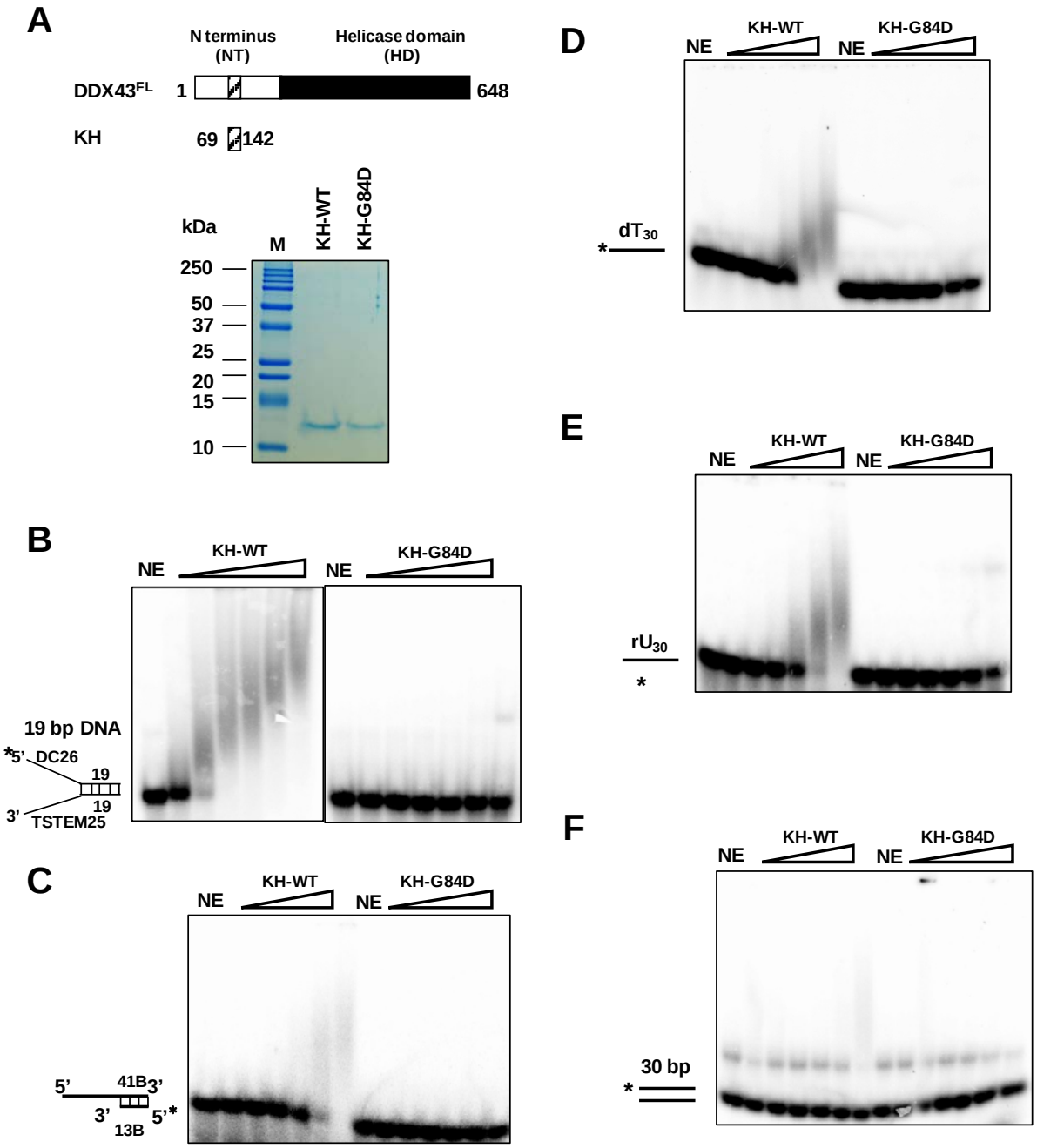


Figure 4.14 Purification and characterization of the DDX43 KH domain. (A) Schematic representation of full-length DDX43 and its KH domain (top), and purified KH domain proteins (wild type and mutant, 1 µg each) shown on Coomassie stained SDS-PAGE gel (bottom). **(B-F)** Representative EMSA images for KH domain proteins binding with 0.5 nM of 19 bp forked duplex DNA **(B)**, 5' tailed dsRNA **(C)** dT30 ssDNA **(D)**, ssRNA rU30 **(E)**, and blunt-end dsDNA **(F)**. The proteins at various concentrations (0.09, 0.18, 0.37, 0.75, 1.5, 3 µM) were incubated with DNA or RNA substrates, and nucleic acids-protein complexes were resolved on native 5% polyacrylamide gels.

4.9.2 The N terminal domain has Strand Exchange Activity

It has been reported that the KH domain-containing fragile X mental retardation protein (FMRP) has strand exchange activity and the presence of KH domains play a critical role for this activity (Gabus *et al*., 2004). Thus, we asked whether DDX43 possesses this nuclear chaperone property or not. Indeed, with unlabeled cold oligo D26 in the reaction the N-terminal domain ($DDX43^{NT}$) displayed strand exchange activity on a 19-bp forked duplex DNA substrate in a concentration (0 to 3 µM) dependent manner, while the N-terminus (1-253aa) mutant G84D ($DDX43^{NT(G84D)}$) did not show this activity **(Fig. 4.15)**.

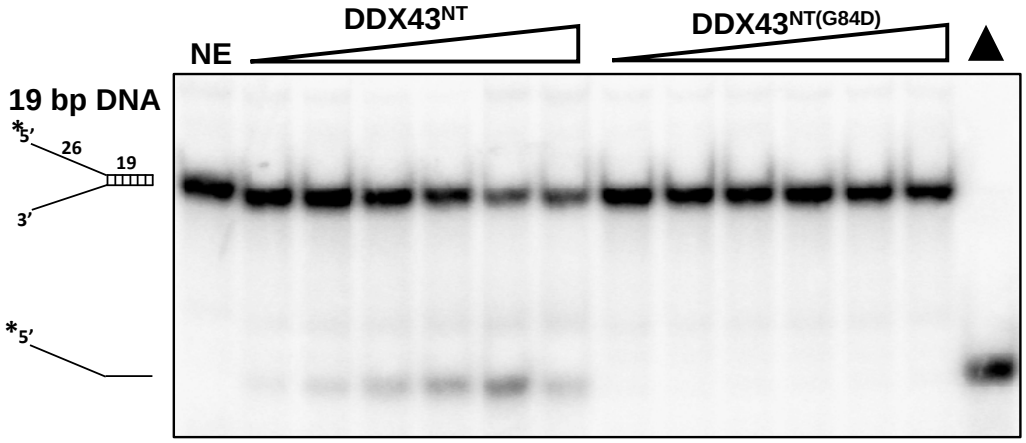


Figure 4.15 Strand exchange activity of DDX43 N-terminal domain. Strand exchange activity was examined by incubating 19 bp forked dsDNA and unlabeled cold oligo with $DDX43^{NT}$ or $DDX43^{NT\ (G84D)}$ (0.09, 0.18, 0.37, 0.75, 1.5, 3 µM). Filled triangle, heat denatured substrate control.

4.9.3 The N-terminal domain is required for the Unwinding Activity

The unwinding activity of the DEAD-box family proteins is only confined to the helicase core. As previous results showed that $DDX43^{NT}$ (1-253aa) of DDX43 possesses strand exchange activity, we aimed to examine if this KH domain containing N-terminus also possesses unwinding activity. Thus, I evaluated the unwinding activity of $DDX43^{NT}$ using a 19bp forked DNA substrate in concentration-dependent manner (0 to 3 µM) and found that it could unwind the substrate but with less efficiency as compare to the DDX43 full length

protein, while mutant DDX43$^{NT(G84D)}$ did not show this activity **(Fig. 4.16 A)**. Furthermore, we found that the unwinding activity of DDX43^{NT} was ATP-dependent **(Fig. 4.16 B)**. In conclusion, I found that the DDX43^{NT} has ATP-dependent strand exchange activity and is required for the full functionality of the protein for unwinding activity.

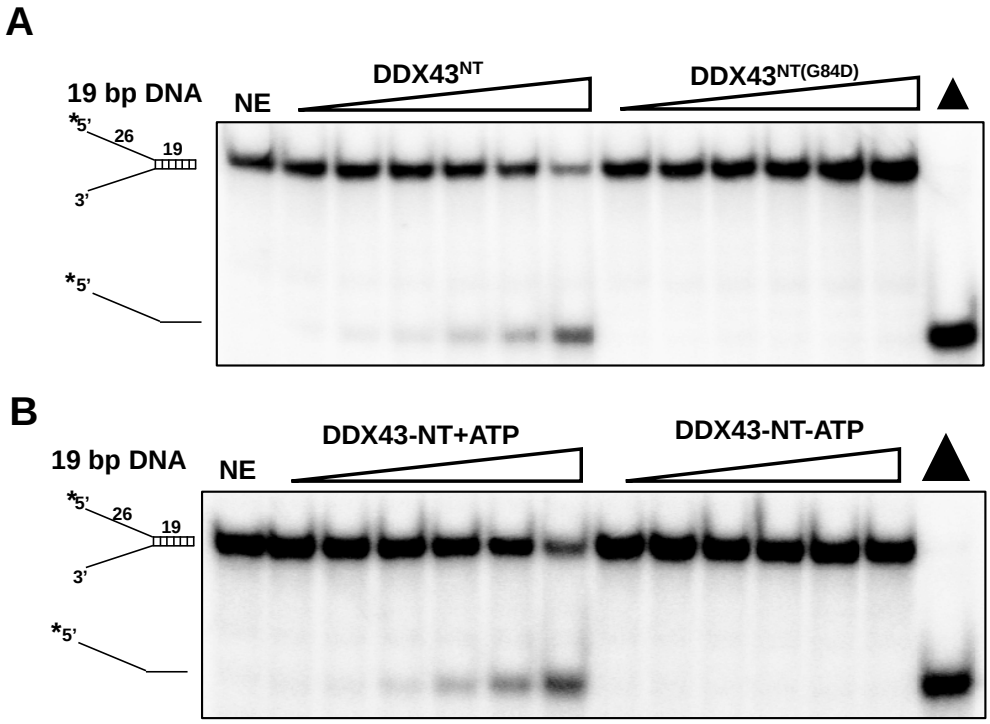


Figure 4.16 Unwinding activity of the DDX43 N-terminal domain. (A-B) Unwinding activity was examined by incubating 19 bp forked dsDNA with DDX43^{NT} or DDX43$^{NT(G84D)}$ (0.09, 0.18, 0.37, 0.75, 1.5, 3 μM). **(A)** Helicase activity was examined by incubating a 19 bp forked dsDNA with DDX43^{NT} or DDX43$^{NT(G84D)}$ (0.09, 0.18, 0.37, 0.75, 1.5, 3 μM) **(B).** Helicase activity of DDX43^{NT} was examined by incubating a 19 bp forked dsDNA with or without ATP. Filled triangle, heat denatured substrate control.

4.10 The N-terminal domain and Helicase Domain Have Synergistic Effects on the Unwinding Activity of DDX43 Protein

4.10.1 The Helicase Domain of DDX43 Possesses Weak Unwinding Activity

Human Pif1 helicase has been shown to contain unwinding activity without its accessory domain (George *et al*., 2009). Next, we asked whether the C-terminal helicase domain alone has unwinding activity and hence I purified the C-terminal helicase domain (254-648 aa, named HD) to near homogeneity (**Fig. 4.17 A).** Using a 19-bp forked duplex DNA substrate,

I detected unwinding for the HD protein in a concentration-dependent manner (0-3μM); however, it was very weak compared to the full-length protein (**Fig. 4.17 B**), suggesting the helicase domain cannot function independently.

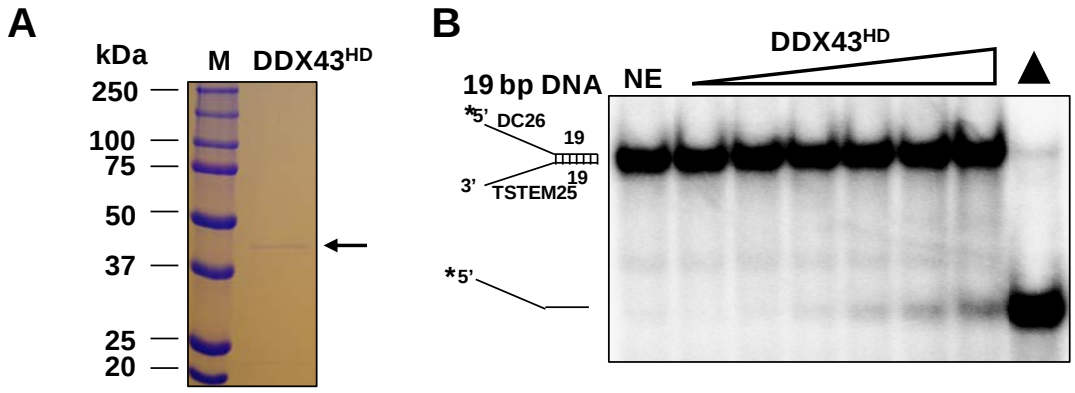


Figure 4.17 Unwinding activity of the DDX43 helicase domain. (**A**) SDS-PAGE analysis of the DDX43 helicase domain proteins eluting from a Sephacryl S-300 HR column. 1 μg was loaded. **(B)** Representative images of helicase reactions by incubating increasing $DDX43^{HD}$ protein (0.09, 0.18, 0.37, 0.75, 1.5, 3 μM) with 0.5 nM a 19 bp forked dsDNA. Filled triangle, heat denatured substrate control.

4.10.2 The N-terminal Domain and Helicase Domain of DDX43 Unwind Substrates Synergistically

Since the DDX43 protein unwinds RNA duplexes with a nonprocessive mechanism and DNA duplexes with a processive mechanism, we suspected that the helicase core domain may be responsible for this phenomenon (Singleton *et al.*, 2007). Then we asked whether the N-terminal domain ($DDX43^{NT}$, 1-253 aa) and helicase domain ($DDX43^{HD}$, 208-648 aa) can work synergistically for unwinding. With an increasing concentration of $DDX43^{NT}$ and $DDX43^{HD}$ proteins (1:1 ratio), we observed increased unwinding on a DNA substrate (**Fig. 4.18 A**); however, it is still lower than the full-length DDX43 protein (**Fig. 4.3 A**). For instance, at the highest concentration (3 μM), the full-length DDX43 protein could unwind ~75%, while the combined $DDX43^{NT}$ and $DDX43^{HD}$ proteins unwound ~50%, with $DDX43^{NT}$ contributing ~40%, and $DDX43^{HD}$ contributing~10% (**Fig. 4.18 B**). Taken

together, these results suggest that the DDX43^{NT} (1-253 aa) binds nucleic acids independent of its C-terminal DDX43^{HD} (254-648 aa), while the C-terminal helicase domain requires N-terminal domain (1-253 aa) for its full activity.

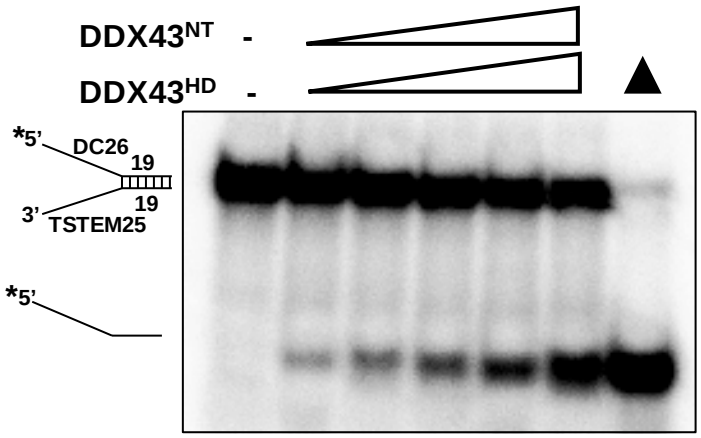


Figure 4.18 Synergistic unwinding activity of the DDX43 helicase domain and the N-terminal domain. Synergic helicase activity of DDX43^{HD} and DDX43^{NT} was examined by incubating 0.5 nM a 19 bp forked dsDNA with both truncated proteins (0.09, 0.18, 0.37, 0.75, 1.5, 3 µM). Filled triangle, heat denatured substrate control.

4.11 The KH Domain of Sam68 Does Not Have a Synergistic Effect with the DDX43 Helicase Domain on Unwinding

4.11.1 The KH domain of Sam68 Protein Binds both DNA and RNA Substrates.

The Src associated substrate in mitosis of 68kDa, Sam68, is an RNA-binding protein which belongs to the STAR (signal transduction and activation of RNA metabolism) family of RNA-binding proteins (Lukong and Richard, 2003). To determine whether the strand exchange and unwinding activity of the N-terminal domain (DDX43^{NT}-1-253aa) from DDX43 are unique, we cloned the KH domain of Sam68^{KH} (97-208 aa) into NdeI/XhoI sites of pET28a vector, overexpressed it in *E. coli.* and purified by two-step chromatography. The Sam68^{KH} protein was purified to near homogeneity **(Fig. 4.19 A)** and an EMSA assay revealed that it bound both 19 bp dsDNA and ssRNA molecules **(Fig. 4.19 B and C)**.

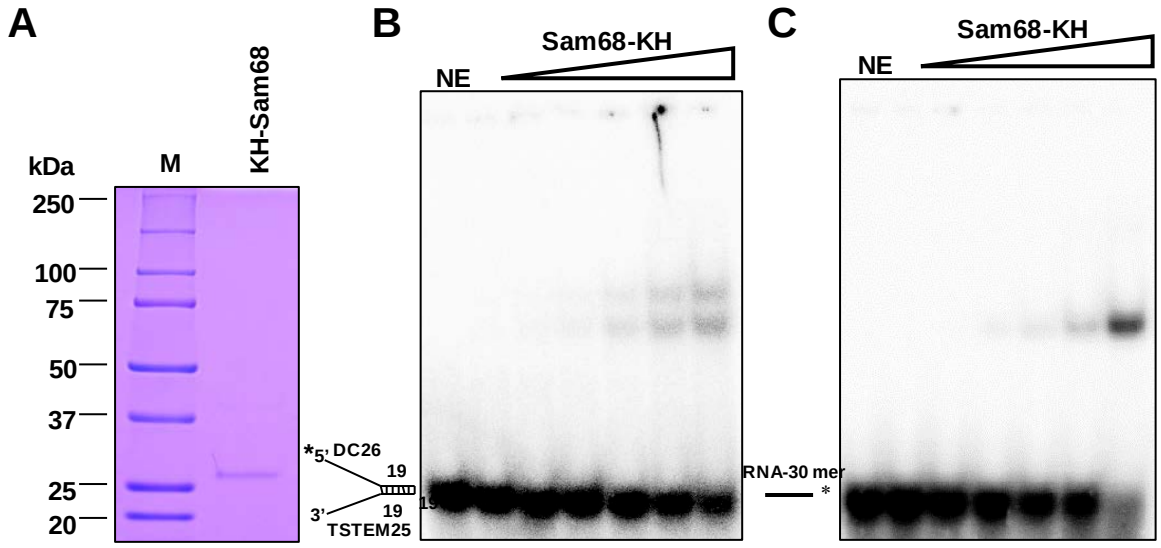


Figure 4.19 Purification and characterization of the KH domain from Sam68. **(A)** SDS-PAGE analysis of Sam68 KH domain protein eluted from a Sephacryl S-100 HR column. **(B,C)** Representative EMSA images for Sam68 KH proteins binding with 0.5 nM 19 bp forked duplex DNA **(B)** and 30 mer random RNA substrate **(C)**. The proteins at various concentrations (0.09, 0.18, 0.37, 0.75, 1.5, 3 µM) were incubated with DNA or RNA substrates, and nucleic acids-protein complexes were resolved on native 5% polyacrylamide gels.

4.11.2 The KH Domain of Sam68 Shows Weak Unwinding on DNA but No Unwinding on an RNA Substrate, and no Synergistic Effect.

As showed in the previous results, DDX43^{NT} (1-253 aa) possesses unwinding activity. Thus next we aimed to examine if an unrelated KH domain in the Sam68 protein (97-208 aa) possesses any helicase activity. Using a 19-bp forked duplex DNA substrate, unexpectedly we found that the KH domain of Sam68 displayed some unwinding activity on DNA substrate **(Fig. 4.20 A)**, though it was less active than DDX43^{NT}. In contrast, Sam68^{KH} exhibited no unwinding activity on a 13 bp duplex RNA substrate **(Fig. 4.20 B)**. Because the DDX43 helicase domain has weak unwinding activity, which is enhanced in the presence of the N-terminal domain (1-253 aa) in a synergistic manner, we asked whether the KH domain of the

N-terminal shows any synergy with the helicase domain of DDX43. However, no such synergy was observed in the presence of the KH domain of Sam68 for unwinding of the substrate **(Fig. 4.20 C)**.

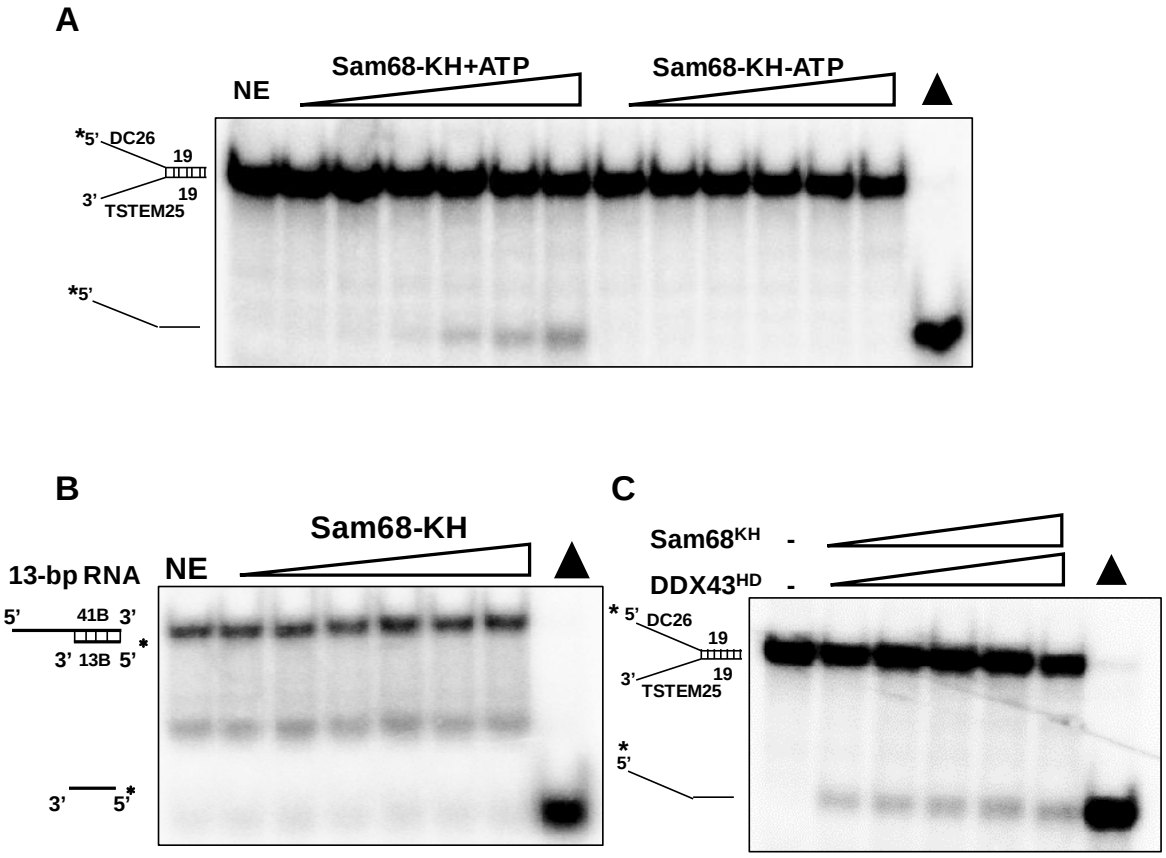


Figure 4.20 Unwinding assays of the Sam68 KH domain protein. **(A-B)** Representative helicase assay for Sam68 KH protein with 19 bp forked duplex DNA with and without ATP **(A)** and 5' tailed 13 bp RNA duplex substrate with ATP **(B)**. **(C)** Helicase activity of $DDX43^{HD}$ and $Sam68^{KH}$ was examined by incubating 0.5 nM 19 bp fork duplex DNA with both proteins (0.09, 0.18, 0.37, 0.75, 1.5, 3 μM). Filled triangle, heat denatured substrate control.

4.12 Overexpression of DDX43 Protein in Triple-Negative Breast Cancer Cell Lines and Tissues

TNBC is a heterogeneous disease, with the poorest prognosis among the different subtypes of breast cancer. At the moment, no standard targeted therapies exist against TNBC. Our immunohistochemistry (IHC) data suggest that the expression of DDX43 varies among tumor

grade and malignancy, but not necessary according to molecular class even though

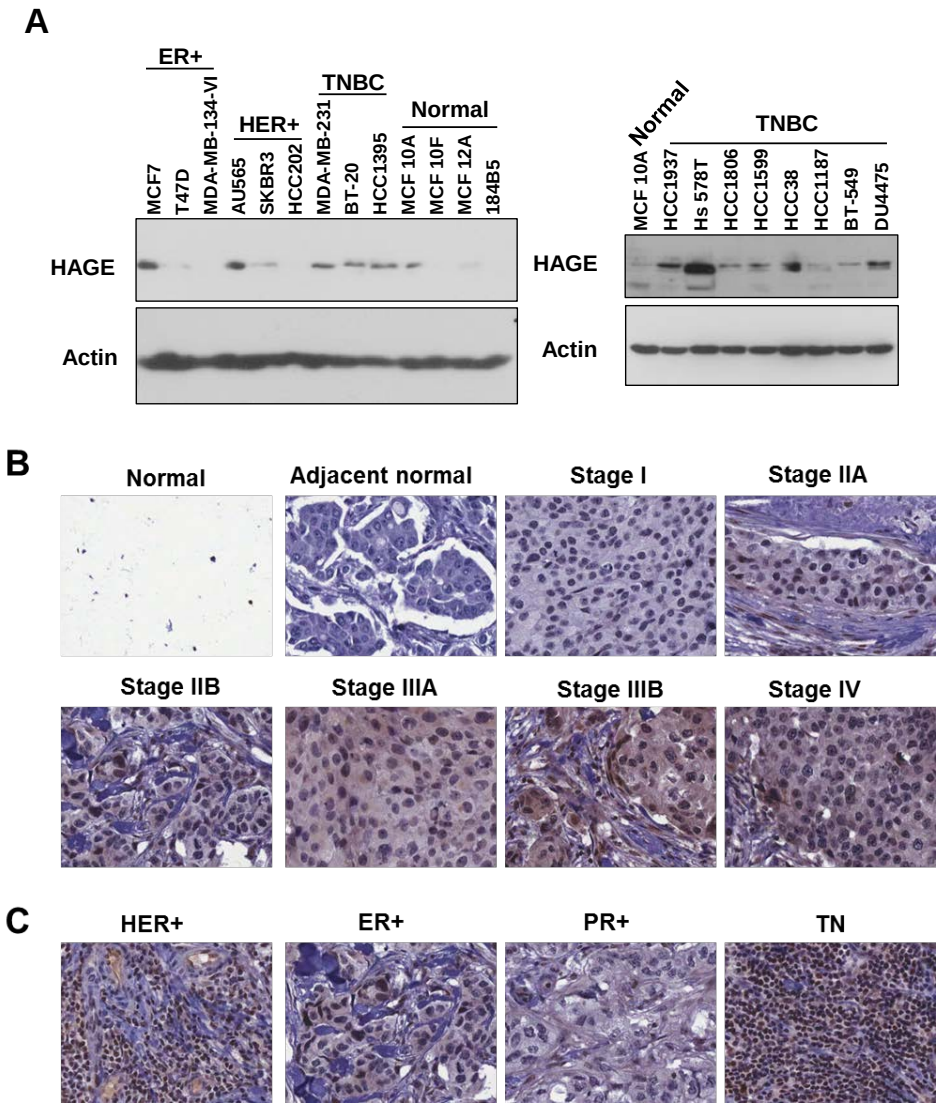


Figure 4.21 The expression of DDX43 in breast cancer cell lines and tissues. (A) Western blot analysis of DDX43 protein expression in indicated breast cell lines. DDX43 antibody is SAB1400618 (Sigma), actin is A5441 (Sigma). A 15-well gel was used for left blot, 10-well for right. **(B)** DDX43 expression in different stages of breast cancer tissues, adjacent cells are used to describe non-invasive breast cancers. Stage I to IV all describe invasive breast cancer. Some of them showed in the figure are divided into subcategories known as IIA, IIB, IIIA and IIIB. Immunohistochemistry was done by US Biomax. **(C)** DDX43 expression in subtypes of breast cancer tissues. Immunohistochemistry was done by US Biomax.

TNBC seems to express higher levels of DDX43. Like primary tumors, there is significant variability amongst breast cancer cell lines based on a number of criteria including gene expression profile, molecular class, subtype, tumorigenicity and metastasis (Neve *et al.*,

2006). To further delineate the significance of DDX43 and triple negative status in breast cancer progression, we evaluated the expression levels among a panel of cell lines representing a spectrum of breast cancer phenotypes. ER+: MCF7, T47D and MDA-MB-134VI, HER+: AU565, SKBR3 and HCC202; TNBC: MDA-MB-231, BT20, and HCC 1395; normal breast cell line: MCF 10A, MCF 10F, MCF 12A, and 184 B5. DDX43 protein expression was detected in MCF 7 of ER+, AU 565 and SKBR3 of HER+, and all three cell lines of TNBC, while low expression in normal cell line MCF 10A and MCF 12A, and absent in MCF 10F and 184 B5 (**Fig. 4.21 A**). These results indicate that DDX43 is consistently expressed in TNBC cell lines. Thus we expanded the TNBC cell line pool, including BT549, HCC 1395, HCC 38, HCC 1806, HCC 1937, HS 578T, and HCC 1187, along with normal cell line MCF 10A, we found that DDX43 is expressed in most of the TNBC. Taken together, we concluded that DDX43 protein is consistently and highly expressed in TNBC cell lines.

4.13 DDX43 Interacts with Proteins that are involved in Pre-mRNA Splicing

To explore the possible biological function of DDX43 in cells, I performed immunoprecipitation by overexpressing DDX43 in HEK-293T cells **(Fig. 4.22 A)** and potential proteins associated with DDX43 were identified by mass spectrometry at the University of Alberta. Our preliminary results suggested that DDX43 interacts with MEP50 and pICln (**Table 4.1**). Their interactions were further confirmed by the presence of pIcln in the HEK293T pull down lysate using antibody against pIcln (1:1000, cat# sc-271454, Santa Cruz) **(Fig. 4.22 B).** MEP50 and pICln are two subunits of the pre-mRNA splicing complex (Friesen *et al*., 2001), thus DDX43 might be involved in pre-mRNA splicing; nevertheless, their physical and functional interaction with DDX43 needs further investigation.

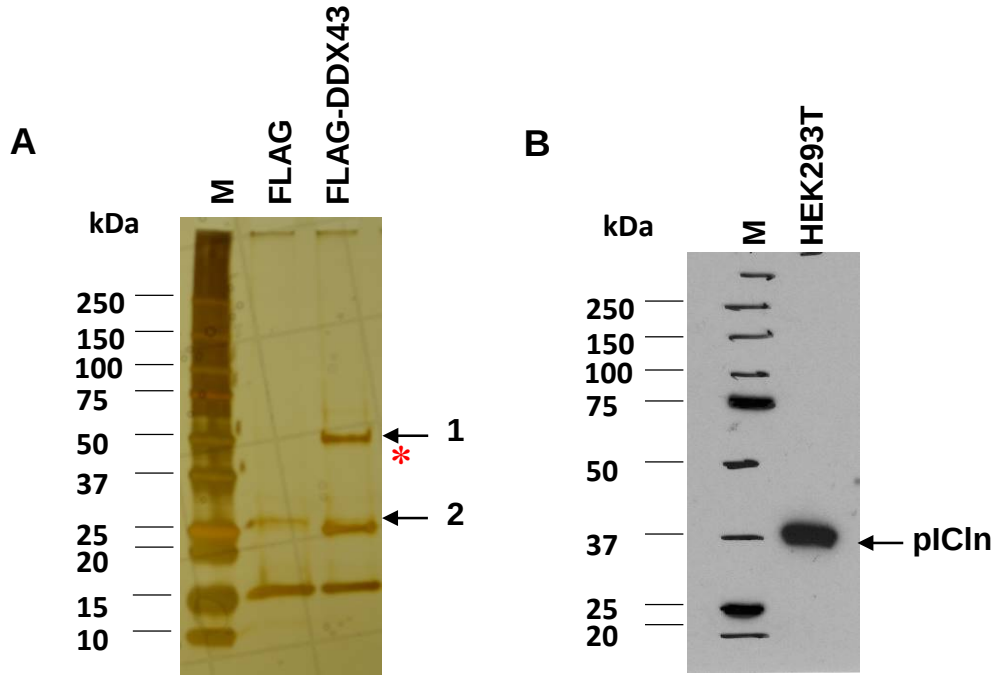


Figure 4.22 Immunoprecipitation of DDX43 and its interacting proteins. (**A**) Silver-stained SDS gel showing overexpressed DDX43 pull-down from HEK293T cells. The Overexpressed DDX43 band is indicated by an asterisk and the potential DDX43-associated protein by arrows. (**B**) Western blot image of pICln (sc-271454, Santa Cruz) using HEK293Tpull down lysate. Note: the predicated size of pICln is 26.2 kDa, but it is about 37 kDa on SDS-PAGE gel.

Table 4.1 Peptide identification by mass spectrometric in-gel digestion analysis

Accession	Description	Score	Coverage	# Unique Peptides	MW
P54105	Methylosome subunit pICln OS=Homo sapiens GN=CLNS1A PE=1 SV=1 - [ICLN_HUMAN]	26.80	15.92	5	26.2
Q9BQA1	Methylosome protein 50 OS=Homo sapiens GN=WDR77 PE=1 SV=1 - [MEP50_HUMAN]	16.62	12.94	4	36.7
P05388	60S acidic ribosomal protein P0 OS=Homo sapiens GN=RPLP0 PE=1 SV=1 - [RLA0_HUMAN]	13.10	13.88	3	34.3
P22626	Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Homo sapiens GN=HNRNPA2B1 PE=1 SV=2 - [ROA2_HUMAN]	10.74	10.55	2	37.4
P29692	Elongation factor 1-delta OS=Homo sapiens GN=EEF1D PE=1 SV=5 - [EF1D_HUMAN]	6.67	17.79	3	31.1
P31942	Heterogeneous nuclear ribonucleoprotein H3 OS=Homo sapiens GN=HNRNPH3 PE=1 SV=2 - [HNRH3_HUMAN]	6.44	10.40	3	36.9
Q9UKM9	RNA-binding protein Raly OS=Homo sapiens GN=RALY PE=1 SV=1 - [RALY_HUMAN]	5.21	5.88	2	32.4
Q96DI7	U5 small nuclear ribonucleoprotein 40 kDa protein OS=Homo sapiens GN=SNRNP40 PE=1 SV=1 - [SNR40_HUMAN]	4.65	4.48	2	39.3
Q8N954	G patch domain-containing protein 11 OS=Homo sapiens GN=GPATCH11 PE=1 SV=3 - [GPT11_HUMAN]	4.64	8.88	2	30.2

5. Discussion

In this study, we have found that DDX43 is an ATP-dependent helicase which could unwind both RNA and DNA substrates, as a dual helicase. It catalyzes the unwinding reaction most efficiently in the presence of Mg^{2+} and ATP. Intriguingly, for unwinding RNA substrates, DDX43 does not follow any strict translocation mechanism, but it unwinds DNA in a unidirectional manner. We also found that the KH domain in the N-terminal region is involved in nucleic acid binding. Surprisingly, the N-terminal domain in the DDX43 exhibits both strand exchange activity and unwinding activity, and synergy between the N-terminal and C-terminal domains of DDX43 is essential for its full unwinding activity. To gain insight into its cellular function, we found that DDX43 interacts with the proteins involved in pre-mRNA splicing, indicating its potential role in RNA metabolism in cells.

5.1 DDX43 is an ATP-dependent Dual Helicase

The majority of known helicases specifically unwind only one type of duplex, either DNA or RNA, but the basis of this discrimination is unknown. DEAD-box family proteins exhibit characteristic RNA helicase activity (Jarmoskaite and Russell, 2011); however, there are some exceptions within the family. For example, the hepatitis C viral NS3 protein is a processive DNA helicase, but can also unwind RNA in the presence of cofactor NS4A (Pang *et al*., 2002). p68, a characteristic RNA helicase also shows unique dual helicase activity (Tuteja *et al*., 2014). Another DEAD-box RNA helicase Dbp9p exhibits DNA-DNA and DNA-RNA helicase activity in the presence of ATP (Kikuma *et al*., 2004). These reports suggest that the DEAD-box family includes proteins with a wide variety of biochemical activities.

DDX43 is classified as a DEAD-box family protein based on sequence homology to the SF2 superfamily. RNA unwinding activity has been reported for DDX43 protein (Linley *et al*., 2012). We started the biochemical characterization of DDX43 protein and the results demonstrated that it could unwind RNA, DNA and RNA:DNA hybrids as well in an ATP-dependent manner. Similar to other DEAD-box RNA helicases, DDX43 displays a limited

processivity on duplex RNA, but a comparatively high processivity on duplex DNA, which makes DDX43 a new entry in the list of dual helicases with unique characteristics.

5.2 DDX43 is a Non-processive RNA Helicase

Processivity is most often defined as the average number of base pairs unwound per helicase binding event, ranging from tens to tens of thousands of base pairs. A highly processive helicase rarely dissociates during unwinding and is thus able to perform many consecutive unwinding steps. A helicase that dissociates more often displays lower processivity and unwinds fewer base-pairs per binding event (Lohman *et al.*, 2008). Because, unlike most helicases, the efficiency of strand separation for these proteins is strongly dependent on the stability of the helix, not just its length.

In general, RNA helicases usually have low processivity on the substrate. Rogers *et al*., showed that eIF4A, a prototypical member of the DEAD-box family, acts in a non-processive manner (Rogers, Jr. *et al*., 1999). They proposed that the initial rate of duplex unwinding and the maximum amount of duplex unwound decreased with increasing stability of the duplex RNA substrate. RhlE has also been found to have low processivity as the ratio (duplex unwound)/ (ATP hydrolyzed) drops 1000-fold upon going from 11 to 14 base pairs (Bizebard *et al*., 2004). Almost all members of the DEAD-box family studied so far appear similar to eIF4A and RhlE, and show little or no processivity. This includes DDX1 (10 bp) (Kellner *et al.*, 2015), DDX20 (15 bp) (Yan *et al*., 2003) and DDX25 (20 bp) (Tang *et al*., 1999). Only p68 (162 bp) and p72 (36 to 46 bp) were shown to unwind moderately long RNA duplexes (Hirling *et al.*, 1989; Rossler *et al*., 2001).

However, it should be noted that reaction conditions, such as the amount of substrate and protein concentration, duration of reaction, amount of ATP, and so on, may cause the difference reported. Even in the case of DDX43, it has been reported it could unwind 40 bp duplex RNA when different reaction conditions were used (Linley, Mathieu *et al*. 2012). Nevertheless, most of the DEAD-box helicases have limited processivity.

In our study, we have found that DDX43 could efficiently unwind 13 bp dsRNA but not 16 bp dsRNA, suggesting that DDX43 is a low processivity DEAD-box protein. It should also be noted that exclusive RNA duplex structures exist mainly during viral replication and transcription. Helical elements within structured RNAs are rarely longer than 10 base pairs (Jankowsky and Fairman, 2007). Thus, there appears to be little need for highly processive RNA helicases.

5.3 DDX43 Unwinds RNA duplexes by a Mode Distinct from Translocating Helicases

Helicases translocate in one direction along the loading strand of DNA or RNA in an ATP-driven manner, thereby displacing either complementary nucleic acids or bound proteins (Mackintosh and Raney, 2006). However, most DEAD-box proteins show bidirectional unwinding. This lack of directionality is unique when compared with other families within SF2 or with other superfamilies. Most DEAD-box proteins unwind substrates without any preference for the polarity of single-stranded overhangs and unwind duplexes by local strand separation.

DEAD-box protein Ded1 has been shown to unwind RNA duplexes irrespective of the orientation of single-stranded regions**,** and it could also unwind blunt-end duplexes although this reaction proceeded at a substantially slower rate (Yang and Jankowsky, 2006). Further additions to this list are DEAD-box RNA helicase proteins p68 (Huang and Liu, 2002), eukaryotic initiation factors eIF-4A (DDX2) (Rozen *et al.*, 1990), DDX25 (Tang *et al.*, 1999), Has1p (Rocak *et al*., 2005), and SRMB (Bizebard *et al*., 2004), which can unwind dsRNA in both 3'-5' and 5'–3' direction, regardless of the orientation of the single-stranded regions. This distinct unwinding mode of DEAD-box proteins appears uniquely suited for the localized separation of short duplexes in the cell. First, usually short duplex RNAs, no longer than 10 bp, are present in cells (Mallam, Del Campo *et al*. 2012). Second, the structure of Mss116p with RNA revealed that a single protein monomer could bind along most of the

length of U_{10} RNA; raising the possibility that strand separation by DEAD-box proteins does not involve any translocation at all (Del Campo and Lambowitz 2009)

In our study, we found that the DDX43 can unwind the blunt end RNA substrate, suggesting that the unwinding could begin internally. However, DDX43 is more active on 5'tailed than 3'tailed dsRNA substrates, but the mechanism remains unknown. A very recent study reported that the Ski2-like RNA helicase Mtr4p unwinds RNA duplexes by 3' to 5' translocation (Patrick *et al*., 2016), suggesting some RNA helicases may unwind RNA duplexes by translocation-based mechanisms. This property of DEAD-box helicases can be related to their role in RNA metabolism as duplex unwinding without translocation on the loading strand may be prevalent in RNA metabolism. These results raise the possibility that many DEAD-box proteins may not function as directional motors, but as "strand separation switches". Unwinding RNA structures by loading the helicase directly onto the duplex and then "switching" the two strands apart, may be well suited for enzymes that encounter mostly short duplexes within large RNP assemblies.

5.4 DDX43 Unwinds a DNA substrate in a 3' to 5' Direction by a Translocation Mechanism

Most of the superfamily2 (SF2) helicases possess 3' to 5' directionality on a DNA substrate (Buttner *et al*., 2007). The conserved domains of SF2 helicases consistently bind nucleic acid strands with a defined polarity which assures the unwinding reaction to be in a 3' to 5' direction. Helicase domain 1 (Rec A1) interacts with the 3' end and helicase domain 2 (Rec A2) interacts with the 5' end of bound single strands, resulting in an interface with a well-defined asymmetry that is likely to differentiate the two possible strand orientations within the binding cleft (Hopfner and Michaelis, 2007). Furthermore, the beta-hairpin that projects from helicase domain 2 of SF2 helicases NS3 and Hel308 have been found to reinforce the directional geometry and stimulates unwinding in 3' to 5' direction (Myong *et al*., 2007). It has been shown that *Plasmodium cynomolgi* DEAD-box DNA helicase 45 (PcDDH45), an

ATP-dependent DNA-unwinding enzyme, translocates in the 3’ to 5’ direction to unwind the substrate. Intriguingly, in our study, we found that DDX43 unwinds the DNA substrate by a conventional canonical mechanism, that is, in a defined 3’-5’ direction. It has been found that the helicase core domains, HD1 and HD2 in Mss116p DEAD box helicase, have modular functions enabling a novel mechanism for RNA-duplex recognition and unwinding where D1 binds ATP and D2 contains a nucleic acid-binding pocket, which accommodates A-form but not B-form duplexes (Mallam *et al*., 2012). Also, the crystal structure of Msp116 with ssDNA and ssRNA revealed that the protein forms a closed-state complex with both nucleic acids. Although the interactions are similar, the closed-state complex with ssRNA contains protein contacts to the RNA 2’-OH groups that are absent in the closed-state complex with ssDNA (Mallam *et al*., 2014). This provides a basis for substrate specificity and might explain the differential unwinding mechanisms on RNA and DNA substrates. Nevertheless, to further address the different unwinding mechanisms of DDX43 on RNA and DNA, we will solve its crystal structure with nucleic acids bound, and perform mutagenesis studies to identify amino acids involved in the polarity.

Most DNA helicases use translocase mechanisms to unwind the substrate, and for translocases processivity is a major aspect of their activity, because their primary function is to walk along nucleic acids. A high degree of processivity is crucial for helicases involved in DNA replication, where millions of base pairs must be replicated quickly (Rajagopal and Patel, 2008). In our study, we found that DDX43 is processive on the DNA substrate compared to that of on the RNA substrate as it could unwind forked dsDNA substrate even up to 40 bp in length. However, the rate of duplex unwinding is significantly decreased with increasing length of the duplex, which is different from classic DNA helicases. A key property of DNA helicases is their ability to unwind long duplexes. Certain DNA helicases (e.g., MCM (Lee and Hurwitz, 2001), RecBCD (Bianco *et al*., 2001), TraI (Lahue and Matson, 1988) can processively unwind DNA tracts with longer than 500 base pairs. Even on a preferred forked duplex DNA substrate, DDX43 acts inefficiently to unwind a 50 bp

duplex, suggesting that DDX43 is specifically tailored to act on short duplex substrates, which might occur in DNA repair or replication restart.

5.5 The KH domain in the N-terminus of DDX43 is Responsible for Nucleic Acid Binding

In eukaryotes, RNA-binding proteins that contain multiple K homology (KH) domains play a key role in coordinating the different steps of RNA synthesis, metabolism and localization (Hollingworth, Candel *et al*. 2012). Both in eukaryotes and prokaryotes KH domains interact with their single-stranded nucleic acids targets with different affinity and specificity and they have been shown to recognize up to four nucleotides, although non-specific contacts can be made with additional flanking nucleotides. KH3 domain in Nova2 has been found to show sequence specific RNA binding (Lewis, Musunuru *et al*. 2000). The highly conserved GXXG loop in the KH domain has a specific sequence element that establishes contact with the nucleic acid, which is essential for the biochemical function of the protein (Valverde *et al.*, 2008). KH domains fold as a three stranded anti-parallel β-sheet on the surface of which pack three α-helices. ssNA binding is mediated by a hydrophobic groove or cleft formed on one side by two short consecutive α -helices (α 1 and α 2) and the intervening GxxG loop and, on the other side, by the inner surface of the domain's β -sheet and the attached variable loop (Musco, Stier *et al*. 1996) **(Fig. 1.4 B)** .

The KH domain has been found in a variety of RNA-binding proteins but, no helicase protein is known to contain this domain to date. KH domains bind RNA or ssDNA, and are found in proteins associated with transcriptional and translational regulation, along with other cellular processes. Several diseases, e.g. fragile X mental retardation syndrome and paraneoplastic disease, are associated with the loss of function of a particular KH domain. Mutation in the hallmark GxxG loop has been found to impair nucleic acid binding of the protein (Hollingworth *et al*., 2012; Chmiel *et al.*, 2006). A single amino acid substitution

(I304N) in this consensus sequence of the FMR1 protein affects its RNA-binding properties and causes fragile X mental-retardation syndrome (De *et al*., 1993).

In our study, sequence alignment revealed that DDX43 contains one KH domain, which has the signature VIGXXGXXI motif in the middle of the domain (**Fig. 4.13 A and B**). Site-directed mutation of glycine at position 84 in the 84-GXXG-87 motif abolished the nucleic acid binding ability of the KH domain. Moreover, a single amino acid mutation $DDX43^{KH(G84D)}$, has significant reduced unwinding activity, indicating the KH domain in DDX43 is responsible for nucleic acid binding.

5.6 The N-terminus of the DDX43 Protein possesses Novel Strand Exchange Activity

Many helicases have been found to possess strand exchange activity including helicases PIF1 (Gu *et al*., 2008) and mitochondrial helicase TWINKLE (Sen *et al*., 2016). Interestingly some non-helicase proteins also possess strand exchange activity as well. The fragile X mental retardation protein has been found to have strand exchange activity and the presence of KH domains play a critical role for this activity (Gabus *et al*., 2004). Our results indicated that N-terminus of DDX43 possesses strand exchange activity and it is require for full functionality of the protein. Because the KH domain is able to bind ssDNA or ssRNA, like ssDNA binding protein RPA and POT1, which have been reported to melt dsDNA and dsRNA (Ahn *et al*., 2009), we believe that the KH domain in DDX43 may have similar duplex destabilizing activity. However, it remains unknown why ATP is required in this reaction. To investigate whether the strand exchange activity is a universal property for KH domain, we are examining the enzymatic activities of KH domain from DDX53 helicase and Sam68 protein.

5.7 Synergistic Effect of the N-terminal domain and the Helicase domain on Unwinding

Several DEAD-box helicases possess ancillary N- and C-terminal domain(s) that are involved in mediating specificity and/or affinity for RNA/DNA substrate (Samatanga and Klostermeier, 2014). It has been reported that DEAD-box helicases bind to helix extensions with low or no specificity, but the ancillary domains help in target recognition and thus unwinding. In DEAD-box proteins DbpA (*E. coli*) and its *B. subtilis* homolog YxiN, the C-terminal domain alone is sufficient to confer specificity for 23S rRNA (Karginov *et al*., 2005). However, for full functionality, it combines the activities of the helicase domain and the C-terminal domain. The C-terminal domain of DbpA recognize a hairpin in the ribosomal 23S rRNA with high affinity and specificity and has been implicated in ribosome biogenesis (Fan *et al*., 2009; Fuller-Pace *et al*., 1993). The helicase core of DEAD-box protein Dbp5/DDX19 is flanked by an N-terminal extension which has been found to contribute to nucleotide binding, a homologous region is also found in DDX25 (Fan *et al*., 2009). The DEAD box protein Hera from *Thermus thermophiles* has been found to contain a flanking C-terminal region to the helicase core, which consists of a dimerization motif and RNA binding domain that mediates high affinity interaction with ribosomal RNA fragments and RNase P RNA (Linden *et al*., 2008).

In our study we have found that the N-terminal domain of DDX43 plays an important role in the full activity of the protein. The C-terminal helicase domain has very low unwinding activity and unwinding activity was significantly increased through synergy between the N-terminal domain and the helicase domain. The truncated protein, i.e. the C-terminal helicase domain, did not show any unwinding activity on RNA and had very weak melting activity on DNA, indicating that the KH domain is more crucial for DDX43 to perform unwinding activity on RNA substrate, which is consistent with the notion that KH domain is a RNA binding domain (Nicastro *et al*., 2015b). More studies are required to determine the structural and functional role of KH domain in DDX43 helicase, for example,

whether the KH domain dictates the sequence specific recognition of substrates for DDX43 helicase, and/or stimulates the unwinding activity of helicase core domain by binding ssRNA/DNA to prevent them reannealing.

5.8 Overexpression of DDX43 in Breast Cancer Cell Lines and Tissues

Breast cancer is the most common and leading cause of cancer deaths in women. Despite the improved outcomes that have been achieved by earlier detection and recent advances in therapy, recurrence still occurs in over 20% of patients (Gluck and Mamounas, 2010); (Dotan and Goldstein, 2010). It is therefore becoming increasingly important to identify biomarkers that can aid in assessing a patient's risk and response to therapy and which might also act as novel targets for immunotherapies (Harris *et al*., 2007; Harbeck *et al*., 2010).

DDX43 was first identified in a sarcoma cell line using representational difference analysis (Martelange *et al*., 2000). DDX43 was subsequently shown to be expressed in many haematological and solid tumour samples but not in any of the normal tissues tested, except the testis (Mathieu *et al*., 2010), thus classified as a Cancer Testis Antigen (CTA). DDX43 has been found as a biomarker for poor prognosis and a predictor of chemotherapy response in breast cancer and its expression is also correlated with the progression of TNBC (Abdel-Fatah *et al*., 2014). It has also been shown that HAGE promotes tumor cell proliferation by regulating NRAS protein expression by unwinding and stabilizing NRAS mRNA and enhancing its transcription, which in turn leads to the activation of downstream signaling pathways AKT phosphorylation through the PI3K pathway and ERK phosphorylation through the RAF/MEK pathway, implicated in cell proliferation and survival (Linley *et al*., 2012b). Thus the tumor-specific expression and oncogenic role of DDX43 in cancer make it a potential target for breast cancer treatment.

In our study, we have found that DDX43 has tumor-specific expression. It expresses minimally in the normal breast cell lines and tissues, and highly in TNBC cell lines and TNBC tissues. Targeted therapies using the overexpression of a specific cancer cell membrane

molecule can facilitate the spatial and temporal delivery of therapeutics (Ovcaricek *et al*., 2011). Although identifying tumour-associated antigens that induce adaptive T-cell immunity is a prerequisite for establishing effective immunotherapy strategies, there is currently a lack of suitable candidates for the development of such treatments in BC. In that scenario, DDX43 can be used as a potential biomarker for treatment of TNBC.

5.9 DDX43 might have a potential role in pre-mRNA splicing

Pre-mRNA splicing is a multistep process that requires two trans-esterification reactions as well as many structural rearrangements of a large ribonucleoprotein complex: the spliceosome. Methylosome subunit pICln is a chaperone that regulates the assembly of spliceosomal U1, U2, U4 and U5 small nuclear ribonucleoproteins (snRNPs), the building blocks of the spliceosome. Thereby, pICln plays an important role in the splicing of cellular pre-mRNAs (Neuenkirchen *et al*., 2008). Methylosome Protein-50 (MEP-50) is a non-catalytic component of the 20S PRMT5-containing methyltransferase complex, which modifies specific arginines to dimethylarginines in several spliceosomal Sm proteins and histones. Using co-immunoprecipitation, our results revealed that both pIcln and MEP50 interact with DDX43, indicating the potential role of DDX43 in the pre-spliceosome assembly of ribonucleoprotein in pre-mRNA splicing. In fact, several DEAD-box RNA helicases have been found to be involved in the early events of the establishment of a functional pre-spliceosome assembly. These helicases have been implicated in RNA secondary structure rearrangements and protein displacement, allowing the spliceosome to assemble, rearrange and dissociate in a controlled manner. In yeast, three DEAD-box proteins have been shown to be required for in vivo splicing including Sub2, Prp28 (DDX23) and Prp5 (DDX46) (Liu and Cheng, 2015). The DEAD-box protein Prp28p appears to function in constitutive and alternative mRNA splicing, and its homolog p72 has been implicated in alternative splicing (Honig *et al*., 2002).

6. Conclusions and Future Work

6.1 Conclusions

In conclusion, we have found that DDX43 is an ATP-dependent dual helicase which can act on both RNA and DNA. It follows translocation mechanism for unwinding DNA substrate; but it unwinds RNA substrate irrespective of the orientation of single-stranded tail. A KH-domain present in the N terminus of DDX43 is involved in nucleic acid binding, and the N-terminal domain possess unique strand exchange activity and helicase activity. C-terminal helicase core domain and N-terminal domain work in a synergistic manner to unwind substrates. DDX43 is highly expressed in both breast cancer tissues and breast cancer cell lines. Interaction of DDX43 with pIcln and MEP-50 indicates its potential physiological functions in RNA metabolism.

6.2. Future Work

1) To determine the functional interaction of DDX43 and associated proteins in pre-mRNA splicing. Three approaches will be used for this experiment;
    a) Co-IP - Using the whole cell lysate with overexpression of FLAG tagged DDX43 in MDA-MB-231 cells, and using DDX43 antibody (HPA031380, Sigma) (Endogenous Co-IP), potentially DDX43 interacting proteins pICln, MEP50, and PRMT5 will be determined.
    b) Immunofluorescence - DDX43, pICln, MEP50, and PRMT5 antibodies will be use to detect their co-localization in HeLa cells under confocal microscope.
    c) Overexpress GFP tagged DDX43 and MCherry tagged pICln, MEP50, or PRMT5 in HEK293T cells, and their co-localization will be detected under confocal microscope.

2) To study the role of DDX43 in oncogenesis, cell proliferation, migration, and tumor invasion will be detected in DDX43-deficient and DDX43-proficient cells. For cell proliferation assays, “soft agar colony formation” will be used using the knockout cell

line and cell number will be determine using Beckman Coulter's cell counter**.** Transwell assay will be used for the cell migration. For cell invasion matrigel will be layered on the top of the well, after adding cell lines on matrigel and chemotactic factors will be added on the bottom and cell invasion will be studied through gel to the well.

3) To examine the RNA splicing defect in breast cancer cell lines and to determine potential pre-mRNA splicing defect in DDX43-deficient cells, DDX43 konckout cells using CRISPR/cas9 technology will be generated. The expression of DDX43 protein will be knock out in TNBC cell lines, including 1) MDA-MB-231, 2) BT20, 3) AU565, and 4) HCC135, and non-tumorigenic cell line MCF-10F will be used as a control.

7. Reference List

Abdel-Fatah, T.M., McArdle, S.E., Johnson, C., Moseley, P.M., Ball, G.R., Pockley, A.G., Ellis,I.O., Rees,R.C., and Chan,S.Y. (2014). HAGE (DDX43) is a biomarker for poor prognosis and a predictor of chemotherapy response in breast cancer. Br. J. Cancer *110*, 2450-2461.

Abdelhaleem, M. (2004). Do human RNA helicases have a role in cancer? Biochim. Biophys. Acta *1704*, 37-46.

Adams, S.P., Sahota, S.S., Mijovic, A., Czepulkowski, B., Padua, R.A., Mufti, G.J., and Guinn, B.A. (2002). Frequent expression of HAGE in presentation chronic myeloid leukaemias. Leukemia *16*, 2238-2242.

Ahn, B., Lee, J.W., Jung, H., Beck, G., and Bohr, V.A. (2009). Mechanism of Werner DNA helicase: POT1 and RPA stimulates WRN to unwind beyond gaps in the translocating strand. PLoS. One. *4*, e4673.

Ambrosini, G., Khanin, R., Carvajal, R.D., and Schwartz, G.K. (2014). Overexpression of DDX43 mediates MEK inhibitor resistance through RAS Upregulation in uveal melanoma cells. Mol. Cancer Ther. *13*, 2073-2080.

Anderson, J.S. and Parker, R.P. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. EMBO J. *17*, 1497-1506.

Ashley, C.T., Jr., Wilkinson, K.D., Reines, D., and Warren, S.T. (1993). FMR1 protein: conserved RNP family domains and selective RNA binding. Science *262*, 563-566.

Baber, J.L., Libutti, D., Levens, D., and Tjandra, N. (1999). High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor. J. Mol. Biol. *289*, 949-962.

Backe, P. H., Messias, A. C, Ravelli, R. B, Sattler, M, Cusack, S .(2005). X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. Structure *13 (7)*: 1055-1067.

Barnes, M., van, R.G., Li, W.M., Mehmood, K., Mackedenski, S., Chan, C.M., King, D.T., Miller, A.L., and Lee, C.H. (2015). Molecular insights into the coding region determinant-binding protein-RNA interaction through site-directed mutagenesis in the heterogeneous nuclear ribonucleoprotein-K-homology domains. J. Biol. Chem. *290*, 625-639.

Bennett, R.J. and Keck, J.L. (2004). Structure and function of RecQ DNA helicases. Crit Rev. Biochem. Mol. Biol. *39*, 79-97.

Bernstein, K.A., Gangloff, S., and Rothstein, R. (2010). The RecQ DNA helicases in DNA repair. Annu. Rev Genet. *44*, 393-417.

Bianco, P.R., Brewer, L.R., Corzett, M., Balhorn, R., Yeh, Y., Kowalczykowski,S.C., and Baskin,R.J. (2001). Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. Nature. *409*, 374-378.

Bianco, P.R. and Webb, M.R. (2012). Helicase unwinding: active or merely perfect? J. Mol. Biol. *420*, 139-140.

Bird, L.E., Subramanya, H.S., and Wigley, D.B. (1998). Helicases: a unifying structural theme? Curr. Opin. Struct. Biol. *8*, 14-18.

Bizebard, T., Ferlenghi, I., Iost, I., and Dreyfus, M. (2004). Studies on three E. coli DEAD-box helicases point to an unwinding mechanism different from that of model DNA helicases. Biochemistry *43*, 7857-7866.

Burd, C.G. and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. Science *265*, 615-621.

Buttner, K., Nehring, S., and Hopfner, K.P. (2007). Structural basis for DNA duplex separation by a superfamily-2 helicase. Nat. Struct. Mol. Biol. *14*, 647-652.

Caruthers, J.M. and McKay, D.B. (2002). Helicase structure and mechanism. Curr. Opin. Struct. Biol. *12*, 123-133.

Causevic, M., Hislop, R.G., Kernohan, N.M., Carey, F.A., Kay, R.A., Steele, R.J., and Fuller-Pace, F.V. (2001). Overexpression and poly-ubiquitylation of the DEAD-box RNA helicase p68 in colorectal tumours. Oncogene *20*, 7734-7743.

Chen, Q., Lin, J., Yao, D.M., Qian, J., Qian, W., Yang, J., Chai, H.Y., Ma, J.C., Deng, Z.Q., Wang, C.Z., and Li, Y. (2012). Aberrant hypomethylation of DDX43 promoter in myelodysplastic syndrome. Br. J. Haematol. *158*, 293-296.

Cheng, W., Brendza, K.M., Gauss, G.H., Korolev, S., Waksman, G., and Lohman, T.M. (2002). The 2B domain of the Escherichia coli Rep protein is not required for DNA helicase activity. Proc. Natl. Acad. Sci. U. S. A *99*, 16006-16011.

Cheng, W., Dumont, S., Tinoco, I., Jr., and Bustamante, C. (2007). NS3 helicase actively separates RNA strands and senses sequence barriers ahead of the opening fork. Proc. Natl. Acad. Sci. U. S. A *104*, 13954-13959.

Chmiel, N.H., Rio, D.C., and Doudna, J.A. (2006). Distinct contributions of KH domains to substrate binding affinity of Drosophila P-element somatic inhibitor protein. RNA. *12*, 283-

291.

Cordin, O., Banroques, J., Tanner, N.K., and Linder, P. (2006). The DEAD-box protein family of RNA helicases. Gene *367*, 17-37.

Daley, J.M., Niu, H., and Sung, P. (2013). Roles of DNA helicases in the mediation and regulation of homologous recombination. Adv. Exp. Med. Biol. *767*, 185-202.

De, B.K., Verkerk, A.J., Reyniers, E., Vits, L., Hendrickx, J., Van, R.B., Van den Bos, F., de, G.E., Oostra, B.A., and Willems, P.J. (1993). A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nat. Genet. *3*, 31-35.

Del Campo, M. and Lambowitz, A.M. (2009). Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. Mol Cell *35*, 598-609.

Donmez, I., Rajagopal, V., Jeong, Y.J., and Patel, S.S. (2007). Nucleic acid unwinding by hepatitis C virus and bacteriophage t7 helicases is sensitive to base pair stability. J. Biol. Chem. *282*, 21116-21123.

Doorbar, J., Elston, R.C., Napthine, S., Raj, K., Medcalf, E., Jackson, D., Coleman, N., Griffin, H.M., Masterson, P., Stacey, S., Mengistu, Y., and Dunlop, J. (2000). The E1E4 protein of human papillomavirus type 16 associates with a putative RNA helicase through sequences in its C terminus. J. Virol. *74*, 10081-10095.

Dotan, E. and Goldstein, L.J. (2010). Optimizing chemotherapy regimens for patients with early-stage breast cancer. Clin. Breast Cancer *10 Suppl 1*, E8-15.

Enemark, E.J. and Joshua-Tor, L. (2008). On helicases and other motor proteins. Curr. Opin. Struct. Biol. *18*, 243-257.

Estruch, F. and Cole, C.N. (2003). An early function during transcription for the yeast mRNA export factor Dbp5p/Rat8p suggested by its genetic and physical interactions with transcription factor IIH components. Mol. Biol. Cell *14*, 1664-1676.

Fairman, M.E., Maroney, P.A., Wang, W., Bowers, H.A., Gollnick, P., Nilsen, T.W., and Jankowsky, E. (2004). Protein displacement by DExH/D "RNA helicases" without duplex unwinding. Science *304*, 730-734.

Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. Curr. Opin. Struct. Biol. *20*, 313-324.

Fan, J.S., Cheng, Z., Zhang, J., Noble, C., Zhou, Z., Song, H., and Yang, D. (2009). Solution and crystal structures of mRNA exporter Dbp5p and its interaction with nucleotides. J. Mol. Biol. *388*, 1-10.

Farah, J.A. and Smith, G.R. (1997). The RecBCD enzyme initiation complex for DNA unwinding: enzyme positioning and DNA opening. J. Mol. Biol. *272*, 699-715.

Flaus, A. and Owen-Hughes, T. (2004). Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? Curr. Opin. Genet. Dev. *14*, 165-173.

Frick, D.N., Banik, S., and Rypma, R.S. (2007). Role of divalent metal cations in ATP hydrolysis catalyzed by the hepatitis C virus NS3 helicase: magnesium provides a bridge for ATP to fuel unwinding. J. Mol. Biol. *365*, 1017-1032.

Friesen, W.J., Paushkin, S., Wyce, A., Massenet, S., Pesiridis, G.S., Van, D.G., Rappsilber, J., Mann, M., and Dreyfuss, G. (2001). The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. Mol. Cell Biol. *21*, 8289-8300.

Fuller-Pace, F.V. (2006). DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. Nucleic Acids Res. *34*, 4206-4215.

Fuller-Pace, F.V. (2013). DEAD box RNA helicase functions in cancer. RNA. Biol. *10*, 121-132.

Fuller-Pace, F.V., Nicol, S.M., Reid, A.D., and Lane, D.P. (1993). DbpA: a DEAD box protein specifically activated by 23s rRNA. EMBO J. *12*, 3619-3626.

Gabus, C., Mazroui, R., Tremblay, S., Khandjian, E.W., and Darlix, J.L. (2004). The fragile X mental retardation protein has nucleic acid chaperone properties. Nucleic Acids Res. *32*, 2129-2137.

Gatfield, D., Le, H.H., Schmitt, C., Braun, I.C., Kocher, T., Wilm, M., and Izaurralde, E. (2001). The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. Curr Biol. *11*, 1716-1721.

George, T., Wen, Q., Griffiths, R., Ganesh, A., Meuth, M., and Sanders, C.M. (2009). Human Pif1 helicase unwinds synthetic DNA structures resembling stalled DNA replication forks. Nucleic Acids Res. *37*, 6491-6502.

Gibson, T.J., Thompson, J.D., and Heringa, J. (1993). The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid. FEBS Lett. *324*, 361-366.

Gilhooly, N.S. and Dillingham, M.S. (2014). Recombination hotspots attenuate the coupled ATPase and translocase activities of an AddAB-type helicase-nuclease. Nucleic Acids Res. *42*, 5633-5643.

Gluck, S. and Mamounas, T. (2010). Improving outcomes in early-stage breast cancer.

Oncology (Williston. Park) *24*, 1-15.

Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., and Blinov, V.M. (1988). A conserved NTP-motif in putative helicases. Nature. *333*, 22.

Grishin, N.V. (2001). KH domain: one motif, two folds. Nucleic Acids Res. *29*, 638-643.

Gu, Y., Masuda, Y., and Kamiya, K. (2008). Biochemical analysis of human PIF1 helicase and functions of its N-terminal domain. Nucleic Acids Res. *36*, 6295-6308.

Guo, M., Hundseth, K., Ding, H., Vidhyasagar, V., Inoue, A., Nguyen, C.H., Zain, R., Lee, J.S., and Wu, Y. (2015). A distinct triplex DNA unwinding activity of ChlR1 helicase. J. Biol. Chem. *290*, 5174-5189.

Hacker, K.J. and Johnson, K.A. (1997). A hexameric helicase encircles one DNA strand and excludes the other during DNA unwinding. Biochemistry *36*, 14080-14087.

Halls, C., Mohr, S., Del, C.M., Yang, Q., Jankowsky, E., and Lambowitz, A.M. (2007). Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. J. Mol. Biol. *365*, 835-855.

Harbeck, N., Salem, M., Nitz, U., Gluz, O., and Liedtke, C. (2010). Personalized treatment of early-stage breast cancer: present concepts and future directions. Cancer Treat. Rev. *36*, 584-594.

Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield,M.R., Hayes,D.F., and Bast,R.C., Jr. (2007). American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J. Clin. Oncol. *25*, 5287-5312.

Hirling, H., Scheffner, M., Restle, T., and Stahl, H. (1989). RNA helicase activity associated with the human p68 protein. Nature *339*, 562-564.

Hollingworth, D., Candel, A.M., Nicastro, G., Martin, S.R., Briata, P., Gherzi, R., and Ramos, A. (2012). KH domains with impaired nucleic acid binding as a tool for functional analysis. Nucleic Acids Res. *40*, 6873-6886.

Honig, A., Auboeuf, D., Parker, M.M., O'Malley, B.W., and Berget, S.M. (2002). Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72. Mol. Cell Biol. *22*, 5698-5707.

Hopfner, K.P. and Michaelis, J. (2007). Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics. Curr. Opin. Struct. Biol. *17*, 87-95.

Huang, Y. and Liu, Z.R. (2002). The ATPase, RNA unwinding, and RNA binding activities of recombinant p68 RNA helicase. J. Biol. Chem. *277*, 12810-12815.

Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. Trends Biochem Sci. *36*, 19-29.

Jankowsky, E. and Bowers, H. (2006). Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. Nucleic Acids Res. *34*, 4181-4188.

Jankowsky, E. and Fairman, M.E. (2007). RNA helicases--one fold for many functions. Curr. Opin. Struct. Biol. *17*, 316-324.

Jarmoskaite, I. and Russell, R. (2011). DEAD-box proteins as RNA helicases and chaperones. Wiley. Interdiscip. Rev. RNA. *2*, 135-152.

Kaplan, D.L., Davey, M.J., and O'Donnell, M. (2003). Mcm4,6,7 uses a "pump in ring" mechanism to unwind DNA by steric exclusion and actively translocate along a duplex. J. Biol. Chem. *278*, 49171-49182.

Karginov, F.V., Caruthers, J.M., Hu, Y., McKay, D.B., and Uhlenbeck, O.C. (2005). YxiN is a modular protein combining a DEx(D/H) core and a specific RNA-binding domain. J. Biol. Chem. *280*, 35499-35505.

Kawaoka, J., Jankowsky, E., and Pyle, A.M. (2004). Backbone tracking by the SF2 helicase NPH-II. Nat. Struct. Mol. Biol. *11*, 526-530.

Kawaoka, J. and Pyle, A.M. (2005). Choosing between DNA and RNA: the polymer specificity of RNA helicase NPH-II. Nucleic Acids Res. *33*, 644-649.

Kellner, J.N., Reinstein, J., and Meinhart, A. (2015). Synergistic effects of ATP and RNA binding to human DEAD-box protein DDX1. Nucleic Acids Res. *43*, 2813-2828.

Kikuma, T., Ohtsu, M., Utsugi, T., Koga, S., Okuhara, K., Eki, T., Fujimori, F., and Murakami, Y. (2004). Dbp9p, a member of the DEAD box protein family, exhibits DNA helicase activity. J. Biol. Chem. *279*, 20692-20698.

Kim, J.L., Morgenstern, K.A., Griffith, J.P., Dwyer, M.D., Thomson, J.A., Murcko, M.A., Lin, C., and Caron, P.R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. Structure. *6*, 89-100.

Lahue, E.E. and Matson, S.W. (1988). *E coli* DNA helicase I catalyzes a unidirectional and highly processive unwinding reaction. J. Biol. Chem. *263*, 3208-3215.

Lee, J.K. and Hurwitz, J. (2001). Processive DNA helicase activity of the minichromosome

maintenance proteins 4, 6, and 7 complex requires forked DNA structures. Proc. Natl. Acad. Sci. U. S. A. *98*, 54-59.

Lee, S., Baek, M., Yang, H., Bang, Y.J., Kim, W.H., Ha, J.H., Kim, D.K., and Jeoung, D.I. (2002). Identification of genes differentially expressed between gastric cancers and normal gastric mucosa with cDNA microarrays. Cancer Lett. *184*, 197-206.

Lewis, H.A., Musunuru, K., Jensen, K.B., Edo, C., Chen, H., Darnell, R.B., and Burley, S.K. (2000). Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. Cell *100*, 323-332.

Li, L., Monckton, E.A., and Godbout, R. (2008). A role for DEAD box 1 at DNA double-strand breaks. Mol. Cell Biol. *28*, 6413-6425.

Liang, X.H. and Fournier, M.J. (2006). The helicase Has1p is required for snoRNA release from pre-rRNA. Mol. Cell Biol. *26*, 7437-7450.

Lin, J., Chen, Q., Yang, J., Qian, J., Deng, Z.Q., Qian, W., Chen, X.X., Ma, J.C., Xiong, D.S., Ma, Y.J., An, C., and Tang, C.Y. (2014). DDX43 promoter is frequently hypomethylated and may predict a favorable outcome in acute myeloid leukemia. Leuk. Res. *38*, 601-607.

Linden, M.H., Hartmann, R.K., and Klostermeier, D. (2008). The putative RNase P motif in the DEAD box helicase Hera is dispensable for efficient interaction with RNA and helicase activity. Nucleic Acids Res. *36*, 5800-5811.

Linder, P. (2003). Yeast RNA helicases of the DEAD-box family involved in translation initiation. Biol. Cell *95*, 157-167.

Linder, P., Lasko, P.F., Ashburner, M., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J., and Slonimski, P.P. (1989). Birth of the D-E-A-D box. Nature *337*, 121-122.

Linley, A.J., Mathieu, M.G., Miles, A.K., Rees, R.C., McArdle, S.E., and Regad, T. (2012). The helicase HAGE expressed by malignant melanoma-initiating cells is required for tumor cell proliferation in vivo. J. Biol. Chem. *287*, 13633-13643.

Liu, Y.C. and Cheng, S.C. (2015). Functional roles of DExD/H-box RNA helicases in Pre-mRNA splicing. J. Biomed. Sci. *22*, 54.

Lohman, T.M., Tomko, E.J., and Wu, C.G. (2008). Non-hexameric DNA helicases and translocases: mechanisms and regulation. Nat. Rev. Mol. Cell Biol. *9*, 391-401.

Lukong, K.E. and Richard, S. (2003). Sam68, the KH domain-containing superSTAR. Biochim. Biophys. Acta *1653*, 73-86.

Mackintosh, S.G. and Raney, K.D. (2006). DNA unwinding and protein displacement by

superfamily 1 and superfamily 2 helicases. Nucleic Acids Res. *34*, 4154-4159.

Mallam, A.L., Del, C.M., Gilman, B., Sidote, D.J., and Lambowitz, A.M. (2012). Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. Nature *490*, 121-125.

Mallam, A.L., Sidote, D.J., and Lambowitz, A.M. (2014). Molecular insights into RNA and DNA helicase evolution from the determinants of specificity for a DEAD-box RNA helicase. Elife. *3*, e04630.

Manosas, M., Xi, X.G., Bensimon, D., and Croquette, V. (2010). Active and passive mechanisms of helicases. Nucleic Acids Res. *38*, 5518-5526.

Martelange, V., De, S.C., De, P.E., Lurquin, C., and Boon, T. (2000). Identification on a human sarcoma of two new genes with tumor-specific expression. Cancer Res. *60*, 3848-3855.

Mathieu, M.G., Knights, A.J., Pawelec, G., Riley, C.L., Wernet, D., Lemonnier, F.A., Straten, P.T., Mueller, L., Rees, R.C., and McArdle, S.E. (2007). HAGE, a cancer/testis antigen with potential for melanoma immunotherapy: identification of several MHC class I/II HAGE-derived immunogenic peptides. Cancer Immunol. Immunother. *56*, 1885-1895.

Mathieu, M.G., Linley, A.J., Reeder, S.P., Badoual, C., Tartour, E., Rees, R.C., and McArdle, S.E. (2010). HAGE, a cancer/testis antigen expressed at the protein level in a variety of cancers. Cancer Immun. *10*, 2.

Mathieu, M.G., Miles, A.K., Ahmad, M., Buczek, M.E., Pockley, A.G., Rees, R.C., and Regad, T. (2014). The helicase HAGE prevents interferon-alpha-induced PML expression in ABCB5+ malignant melanoma-initiating cells by promoting the expression of SOCS1. Cell Death. Dis. *5*, e1061.

Matunis, E.L., Matunis, M.J., and Dreyfuss, G. (1992). Characterization of the major hnRNP proteins from Drosophila melanogaster. J. Cell Biol. *116*, 257-269.

McGeoch, A.T., Trakselis, M.A., Laskey, R.A., and Bell, S.D. (2005). Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. Nat. Struct. Mol. Biol. *12*, 756-762.

Meetei, A.R., Medhurst, A.L., Ling, C., Xue, Y., Singh, T.R., Bier, P., Steltenpool, J., Stone, S., Dokal, I., Mathew, C.G., Hoatlin, M., Joenje, H., de Winter, J.P., and Wang, W. (2005). A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. Nat. Genet. *37*, 958-963.

Mercier, E., Girodat, D., and Wieden, H.J. (2015). A conserved P-loop anchor limits the

structural dynamics that mediate nucleotide dissociation in EF-Tu. Sci. Rep. *5*, 7677.

Musco, G., Stier, G, Joseph, C, Castiglione Morelli, M. A, Nilges, M, Gibson, T. J, Pastore, A.. (1996). Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. Cell. *85*, 237-245.

Myong, S., Bruno, M.M., Pyle, A.M., and Ha, T. (2007). Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. Science. *317*, 513-516.

Neuenkirchen, N., Chari, A., and Fischer, U. (2008). Deciphering the assembly pathway of Sm-class U snRNPs. FEBS Lett. *582*, 1997-2003.

Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.P., Tong, F., Speed, T., Spellman, P.T., DeVries, S., Lapuk, A., Wang, N.J., Kuo, W.L., Stilwell, J.L., Pinkel, D., Albertson, D.G., Waldman, F.M., McCormick, F., Dickson, R.B., Johnson, M.D., Lippman, M., Ethier, S., Gazdar, A., and Gray, J.W. (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell *10*, 515-527.

Nicastro, G., Taylor, I.A., and Ramos, A. (2015). KH-RNA interactions: back in the groove. Curr. Opin. Struct. Biol. *30*, 63-70.

Nonin, S., Leroy, J.L., and Gueron, M. (1995). Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying. Biochemistry. *34*, 10652-10659.

Ovcaricek, T., Frkovic, S.G., Matos, E., Mozina, B., and Borstnar, S. (2011). Triple negative breast cancer - prognostic factors and survival. Radiol. Oncol. *45*, 46-52.

Pang, P.S., Jankowsky, E., Planet, P.J., and Pyle, A.M. (2002). The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. EMBO J. *21*, 1168-1176.

Patel, S.S. and Donmez, I. (2006). Mechanisms of helicases. J. Biol. Chem. *281*, 18265-18268.

Patel, S.S. and Picha, K.M. (2000). Structure and function of hexameric helicases. Annu. Rev Biochem. *69*, 651-697.

Patrick, E.M., Srinivasan, S., Jankowsky, E., and Comstock, M.J. (2016). The RNA helicase Mtr4p is a duplex-sensing translocase. Nat. Chem. Biol. *13*, 99-104

Pause, A., Methot, N., and Sonenberg, N. (1993). The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP

hydrolysis. Mol. Cell Biol. *13*, 6789-6798.

Putnam, C.D., Clancy, S.B., Tsuruta, H., Gonzalez, S., Wetmur, J.G., and Tainer, J.A. (2001). Structure and mechanism of the RuvB Holliday junction branch migration motor. J. Mol. Biol. *311*, 297-310.

Pyle, A.M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. Annu. Rev. Biophys. *37*, 317-336.

Rajagopal, V. and Patel, S.S. (2008). Single strand binding proteins increase the processivity of DNA unwinding by the hepatitis C virus helicase. J. Mol. Biol. *376*, 69-79.

Robert, F. and Pelletier, J. (2013). Perturbations of RNA helicases in cancer. Wiley. Interdiscip. Rev. RNA. *4*, 333-349.

Rocak, S., Emery, B., Tanner, N.K., and Linder, P. (2005). Characterization of the ATPase and unwinding activities of the yeast DEAD-box protein Has1p and the analysis of the roles of the conserved motifs. Nucleic Acids Res. *33*, 999-1009.

Rogers, G.W., Jr., Richter, N.J., and Merrick, W.C. (1999). Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. J. Biol. Chem. *274*, 12236-12244.

Roman-Gomez, J., Jimenez-Velasco, A., Agirre, X., Castillejo, J.A., Navarro, G., San Jose-Eneriz, E., Garate, L., Cordeu, L., Cervantes, F., Prosper, F., Heiniger, A., and Torres, A. (2007). Epigenetic regulation of human cancer/testis antigen gene, HAGE, in chronic myeloid leukemia. Haematologica *92*, 153-162.

Rossler, O.G., Straka, A., and Stahl, H. (2001). Rearrangement of structured RNA via branch migration structures catalysed by the highly related DEAD-box proteins p68 and p72. Nucleic Acids Res. *29*, 2088-2096.

Rozen, F., Edery, I., Meerovitch, K., Dever, T.E., Merrick, W.C., and Sonenberg, N. (1990). Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. Mol. Cell Biol. *10*, 1134-1144.

Samatanga, B. and Klostermeier, D. (2014). DEAD-box RNA helicase domains exhibit a continuum between complete functional independence and high thermodynamic coupling in nucleotide and RNA duplex recognition. Nucleic Acids Res. *42*, 10644-10654.

Schmidt, U., Lehmann, K., and Stahl, U. (2002). A novel mitochondrial DEAD box protein (Mrh4) required for maintenance of mtDNA in Saccharomyces cerevisiae. FEMS Yeast Res. *2*, 267-276.

Sen, D., Patel, G., and Patel, S.S. (2016). Homologous DNA strand exchange activity of the human mitochondrial DNA helicase TWINKLE. Nucleic Acids Res. *44*, 4200–4210.

Shibuya, T., Tange, T.O., Sonenberg, N., and Moore, M.J. (2004). eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. Nat. Struct. Mol. Biol. *11*, 346-351.

Silverman, E., Edwalds-Gilbert, G., and Lin, R.J. (2003). DExD/H-box proteins and their partners: helping RNA helicases unwind. Gene *312*, 1-16.

Singh, D.K., Ghosh, A.K., Croteau, D.L., and Bohr, V.A. (2012). RecQ helicases in DNA double strand break repair and telomere maintenance. Mutat. Res. *736*, 15-24.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annu. Rev. Biochem. *76*, 23-50.

Singleton, M.R. and Wigley, D.B. (2002). Modularity and specialization in superfamily 1 and 2 helicases. J. Bacteriol. *184*, 1819-1826.

Siomi, H., Choi, M., Siomi, M.C., Nussbaum, R.L., and Dreyfuss, G. (1994). Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. Cell. *77*, 33-39.

Solem, A., Zingler, N., and Pyle, A.M. (2006). A DEAD protein that activates intron self-splicing without unwinding RNA. Mol. Cell *24*, 611-617.

Sommers, J.A., Rawtani, N., Gupta, R., Bugreev, D.V., Mazin, A.V., Cantor, S.B., and Brosh, R.M., Jr. (2009). FANCJ uses its motor ATPase to destabilize protein-DNA complexes, unwind triplexes, and inhibit RAD51 strand exchange. J. Biol. Chem. *284*, 7505-7517.

Soultanas, P. and Wigley, D.B. (2001). Unwinding the 'Gordian knot' of helicase action. Trends Biochem. Sci. *26*, 47-54.

Spingola, M., Armisen, J., and Ares, M., Jr. (2004). Mer1p is a modular splicing factor whose function depends on the conserved U2 snRNP protein Snu17p. Nucleic Acids Res. *32*, 1242-1250.

Story, R.M., Li, H., and Abelson, J.N. (2001). Crystal structure of a DEAD box protein from the hyperthermophile Methanococcus jannaschii. Proc. Natl. Acad. Sci. U. S. A. *98*, 1465-1470.

Subramanya, H.S., Bird, L.E., Brannigan, J.A., and Wigley, D.B. (1996). Crystal structure of a DExx box DNA helicase. Nature. *384*, 379-383.

Suhasini, A.N. and Brosh, R.M., Jr. (2013). DNA helicases associated with genetic instability,

cancer, and aging. Adv. Exp. Med. Biol. *767*, 123-144.

Tang, P.Z., Tsai-Morris, C.H., and Dufau, M.L. (1999). A novel gonadotropin-regulated testicular RNA helicase. A new member of the dead-box family. J. Biol. Chem. *274*, 37932-37940.

Taylor, S.D., Solem, A., Kawaoka, J., and Pyle, A.M. (2010). The NPH-II helicase displays efficient DNA x RNA helicase activity and a pronounced purine sequence bias. J. Biol. Chem. *285*, 11692-11703.

Tuteja, N., Tarique, M., Banu, M.S., Ahmad, M., and Tuteja, R. (2014). Pisum sativum p68 DEAD-box protein is ATP-dependent RNA helicase and unique bipolar DNA helicase. Plant Mol. Biol. *85*, 639-651.

Uhlmann-Schiffler, H., Jalal, C., and Stahl, H. (2006). Ddx42p--a human DEAD box protein with RNA chaperone activities. Nucleic Acids Res. *34*, 10-22.

Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., Zeeberg, B.R., Kane, D., Weinstein, J.N., Blume, J., and Darnell, R.B. (2005). Nova regulates brain-specific splicing to shape the synapse. Nat. Genet. *37*, 844-852.

Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. FEBS J. *275*, 2712-2726.

Valverde, R., Pozdnyakova, I., Kajander, T., Venkatraman, J., and Regan, L. (2007). Fragile X mental retardation syndrome: structure of the KH1-KH2 domains of fragile X mental retardation protein. Structure. *15*, 1090-1098.

von Hippel, P.H. and Delagoutte, E. (2001). A general model for nucleic acid helicases and their "coupling" within macromolecular machines. Cell. *104*, 177-190.

Wei, X., Pacyna-Gengelbach, M., Schluns, K., An, Q., Gao, Y., Cheng, S., and Petersen, I. (2004). Analysis of the RNA helicase A gene in human lung cancer. Oncol. Rep. *11*, 253-258.

Wiese, M. and Pajeva, I.K. (2014). HAGE, the helicase antigen as a biomarker for breast cancer prognosis (WO2013144616). Expert. Opin. Ther. Pat. *24*, 723-725.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. Nature. *407*, 327-339.

Wockner, L.F., Morris, C.P., Noble, E.P., Lawford, B.R., Whitehall, V.L., Young, R.M., and Voisey, J. (2015). Brain-specific epigenetic markers of schizophrenia. Transl. Psychiatry. *5*, e680.

Wu, Y. and Brosh, R.M., Jr. (2012). DNA helicase and helicase-nuclease enzymes with a conserved iron-sulfur cluster. Nucleic Acids Res. *40*, 4247-4260.

Xu, Y.Z., Newnham, C.M., Kameoka, S., Huang, T., Konarska, M.M., and Query, C.C. (2004). Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. EMBO J. *23*, 376-385.

Yan, X., Mouillet, J.F., Ou, Q., and Sadovsky, Y. (2003). A novel domain within the DEAD-box protein DP103 is essential for transcriptional repression and helicase activity. Mol. Cell Biol. *23*, 414-423.

Yang, Q., Del, C.M., Lambowitz, A.M., and Jankowsky, E. (2007). DEAD-box proteins unwind duplexes by local strand separation. Mol. Cell. *28*, 253-263.

Yang, Q. and Jankowsky, E. (2005). ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. Biochemistry. *44*, 13591-13601.

Yang, Q. and Jankowsky, E. (2006). The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. Nat. Struct. Mol. Biol. *13*, 981-986.

Zhou, Y., Mah, T.F., Greenblatt, J., and Friedman, D.I. (2002). Evidence that the KH RNA-binding domains influence the action of the E. coli NusA protein. J. Mol. Biol. *318*, 1175-1188.